Scanning the medical phenome to identify new medical diagnoses after recovery from COVID-19 in a US cohort

Supplementary Appendix

June 5, 2022

Table of Contents

Supplementary Methods

A. Study setting

Study participants came from Vanderbilt University Medical Center (VUMC), a private nonprofit academic medical institution based in Nashville, Tennessee, USA. It is one of the largest medical centers in the southeastern United States and serves as an anchor for specialty and primary care for patients throughout Tennessee and the Mid-South region. VUMC has maintained an EHR since the mid-1990's and largely eliminated paper records in clinical care since 2004.[1] In 2019 before the start of the pandemic, VUMC had 1,131 licensed beds, and managed over 2 million ambulatory visits, over 110,000 emergency department visits, and over 55,000 surgical procedures annually.

B. Vanderbilt Research Derivative (RD)

The Research Derivative is a database of clinical and related data derived from the Vanderbilt University Medical Center's (VUMC) clinical systems and restructured for research.[1] Data is repurposed from VUMC's enterprise data warehouse (EDW), which includes data from clinical information systems eStar (VUMC's local implementation of Epic Hyperspace®), VPIMS (Vanderbilt Perioperative Information Management Systems), ORMIS (Operating Room Management Information System), and clinical laboratory systems, as well as legacy systems including StarPanel (Vanderbilt's native electronic medical records system), Horizon Export Orders, and others. Data is transformed from the EDW to the RD using a set of custom Extract, Transform, and Load (ETL) pipelines to map data to the Observational Medical Outcomes Partnership (OMOP) common data model.[2] The medical record number and other person identifiers are preserved within the database. Data types include reimbursement codes, clinical notes and documentation, nursing records, medication data, laboratory data, encounter and visit data, and respiratory flowsheet data on mechanical ventilator and oxygen usage. Output may include structured data points, such as International Classification of Diseases (ICD) codes and encounter dates, semi-structured data such as laboratory tests and results, or unstructured data such as physician progress reports. Pertinent to this study, laboratory tests are coded as Logical Observation Identifiers Names and Codes (LOINC) and diagnoses are coded using the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM). The RD database is stored on a secure database server housed in the Vanderbilt Data Center. The database is fully compliant with the administrative, physical, and technical provisions of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Security and Privacy Rules, and operates with

oversight from the Vanderbilt Institutional Review Board. The database is maintained by VUMC's Office of Research Informatics and the Vanderbilt Institute for Clinical and Translational Research (VICTR) Big Data team under the direction of Paul Harris, PhD.

C. Vanderbilt COVID-19 registry database

The COVID-19 registry utilizes a customized version of the RD specifically created to support COVID-related research efforts.[3] It captures all patients who had SARS-CoV-2 testing performed at VUMC and its associated community testing sites, outpatient clinics, urgent care centers, and affiliate hospitals. The registry collects a broad range of health data including demographics, vital signs, medical and social history, medications, self-reported symptoms, visit information, clinic notes, respiratory flowsheets, diagnostic codes, and procedural codes. The COVID-19 registry is refreshed daily to facilitate close to real-time access to research data, rather than the usual monthly refresh for the "traditional" RD as described above. Additional COVID-19-specific tables are included in the registry database specifically related to SARS-CoV-2 testing including testing dates, care sites, ordering provider, test status, indication for testing (symptoms consistent with clinical COVID-19, or one of several institutionally-determined acceptable indications for asymptomatic testing), along with test results.

D. Investigator access to database population

The study authors had full access to all patient and clinical data available in the RD and COVID-19 registry database.

E. Quality control of study data

Patient-level quality control. We excluded nine patients with missing data on date of birth or sex, or had zero visit records present in the EHR (ire. even no visit associated with a SARS-CoV-2 PCR test). We included patients who had race or ethnicity coded as "Unknown / Not reported" as some patients choose not to provide a self-identified race or ethnicity when being registered for a care visit.

SARS-CoV-2 testing data quality control. we identified the records of all SARS-CoV-2 PCR tests included in VUMC's research derivative using the LOINC code 94533-7. We excluded all tests that were performed after the data censoring date (January 1, 2022) and those with invalid dates or clearly erroneous dates (e.g. testing dates before February 1, 2020). We only included test results which were

identified as completed, and not still in progress or never resulted. We also only included those laboratory test records which had a result of either "Positive" or "Not Detected", and excluded any PCR tests with indeterminate, pending, or canceled results.

ICD code and phecode data quality control. As the COVID-19 registry database is updated daily, some data is subject to change as they are finalized, amended, corrected, or updated in the medical record. Pertinent to this study, ICD diagnosis codes are typically entered into the enterprise data warehouse by practitioners, but can be changed during finalization by coding specialists several days or weeks later. Our experience and that of the VICTR Big Data staff has been that ICD codes and problem lists have relatively low volatility over time.[1] To mitigate the potential for any changes in diagnosis codes, we extracted the diagnosis codes from the registry database in late February 2022 which was several weeks after the study's data censoring date.

Vital sign and laboratory value data quality control. We captured vital signs and clinical laboratory test results (**Table S2**) for each patient obtained during outpatient encounters for two separate time period: (1) results within 180 days before the index SARS-CoV-2 and (2) results within 365 days after the recovery date. We took the median value for each period for final analyses. Similar to our experience with diagnosis codes, we have found relatively low volatility for these results once they are finalized in the EHR.[1]

F. PheWAS model design and covariates

Exclusions from specific phenotype analyses. As described in the **Methods** section of the main manuscript, for each phenotype we excluded any patient who had the corresponding phecode in their medical record prior to reaching the recovery period. Reasons for this included (1) diagnoses entered into the EHR prior to the index SARS-CoV-2 PCR test, (2) diagnoses made during an inpatient hospitalization associated with the index PCR test, or (3) diagnoses made before the patient reached the 30-day recovery period. We allowed an unlimited look-back period to identify phecodes entered prior to SARS-CoV-2 testing to minimize the risk of misclassifying old diagnoses as "new" in the post-acute period. The electronic health record at Vanderbilt dates back to the early 1990s,[4] but in practice the overwhelming majority of pre-testing phecodes (98.9%) in our cohort were for diagnosis occurring within the last 20 years (after 2001).

Phecode analyses: For PheWAS analyses using phecodes, we required a minimum of at least 10 phenotype cases in the overall cohort. Covariates included in the PheWAS models included age, sex,

race (white, black, other/multi-ethnic, or not reported), ethnicity (Hispanic/Latino, Non-Hispanic/Non-Latino, or not reported), time under observation in the post-acute period (days), testing indication (symptomatic or asymptomatic testing), and Charlson index comorbidities present prior to SARS-CoV-2 testing using a phecode-based definition (**I. Model Variables; Supplementary Table S2**).[5,6]

Vital sign / clinical laboratory test analyses: Covariates included in linear regression models evaluating changes in vital sign or clinical laboratory tests included age, sex, race, ethnicity, and time between the pre-testing value and the post-acute phase value (in days).

G. Loss of follow-up and missing phecode data

All patients who underwent SARS-CoV-2 testing were included in the primary analysis regardless whether they had a subsequent visit at VUMC. This is in keeping with usual practice for PheWAS as the methodology assumes that for each phenotype patients without relevant diagnosis code did not develop the phenotype and thus can be considered controls.[7]

H. Sensitivity analyses

We assessed the robustness of our findings to our model assumptions using several sensitivity analyses. Firstly, we assessed the effect of our control definition and loss to follow-up by limiting the analysis to those patients who had at least one follow-up visit during the post-acute phase. Secondly, we assessed the effect of patients newly seeking care in our system during the pandemic by evaluating only those patients who had at least 2 (two) visits in our system separated by at least 6 months prior to the start of the pandemic. Thirdly, we assessed the effects of the standard PheWAS "phenotype case" by using a more relaxed "phenotype case" definition requiring a phecode only on a single visit. Lastly, we assessed the effect of bias from differences in baseline clinical variables using a propensity-matched sub-cohort which matched 3 never-infected controls to each COVID-19 survivor. Further details are provided in **Section J**.

I. PheWAS model variables

1) **Age** at time of testing (years)
2) **Sex** coded as female or male.
3) **Race** coded as 4 separate binary variables: White, Black or African American, Other race or multi-racial, or Unknown/not reported.
4) **Ethnicity** coded as 3 separate binary variables: Hispanic or Latino, Non-Hispanic/Non-Latino, or Unknown/not reported.
5) **Post-acute observation time**. Number of days from start of post-acute phase (as defined in **Methods** of the main manuscript) until date of data censoring.
6) **Asymptomatic testing**. Binary variable indicating the patient's SARS-CoV-2 test was performed for asymptomatic screening versus symptomatic testing. Reasons for asymptomatic screening included asymptomatic admission to the hospital for another diagnosis, pre-procedural or pre-surgical screening, known SARS-CoV-2 exposure, pre-receipt of immunosupressive or anti-neoplastic therapy, pre-transplant evaluation, or requirement for placement in post-acute care or long-term nursing care.
7) **Inpatient hospitalization**. Binary variable indicating if the SARS-CoV-2 test was performed within 15 days prior to an inpatient hospitalization, or performed during an inpatient hospitalization.
8) **Charlson comorbidities**. Each comorbidity is encoded as a binary variable indicating presence of corresponding phecode(s) on one or more visits at least 15 days prior to SARS-CoV-2 test (See **Table S2**).
   i. **Myocardial infarction**.
   ii. **Congestive heart failure**
   iii. **Peripheral vascular disease**
   iv. **Cerebrovascular disease**
   v. **Dementia**
   vi. **Chronic pulmonary disease**
   vii. **Rheumatologic disease**
   viii. **Peptic ulcer disease**
   ix. **Diabetes**
   x. **Mild liver disease**
   xi. **Severe liver disease**
   xii. **Hemiplegia or paraplegia**
   xiii. **Renal disease**
   xiv. **Any malignancy, including lymphoma or leukemia, not non-melanoma skin cancer.**
   xv. **Metastatic solid tumor**
   xvi. **AIDS or HIV infection**

J. Propensity matching analysis

As imbalances in some clinical variables were noted between the SARS-CoV-2 positive and never-infected groups, we compared the results of performing a PheWAS using the full (unmatched) study cohort with results using a propensity-score matched sub-cohort. We generated a propensity score to estimate probability of having a positive SARS-CoV-2 test using a generalized linear model with a probit link function. We conditioned the propensity scoring model on age, sex, race, ethnicity, symptomatic testing indication, inpatient hospitalization around time of SARS-CoV-2 test, observation time after recovery (in days), and length of EHR prior to SARS-CoV-2 testing (in years). We then performed nearest neighbor matching without replacement with a 3:1 control-case ratio. Control-case ratios of 1:1 and 2:1 were also assessed but did not result in satisfactory matching performance. After matching, all standardized mean differences in the conditioning variables were below 0.1 and Kolmogorov-Smirnov statistics (maximum difference in empirical cumulative density function) demonstrated improvement for all variables (**Figure S8**), indicating acceptable matching between cases and controls (**Table S13**). Visual assessment of continuously distributed conditioning variables (age, observation time after recovery, and EHR length prior to SARS-CoV-2 testing) demonstrated that the matching improved alignment between the SARS-CoV-2 positive and never-infected control group for all variable (**Figure S9**). We did note a modest persistent difference in distributions of observation time after recovery between groups, with the SARS-CoV-2 group having a more peaked distribution of post-recovery observation time versus a more uniform distribution among the never-infected controls (**Figure S9, B**). This reflected the higher positive test rates during local waves of the pandemic, whereas the distribution of negative tests was spread more uniformly over time. Noting this modest persistent difference between exposure groups for post-recovery observation time, post-matching standardized mean difference for this variable remained acceptable at 0.045 (**Table S13**). Matching was performed using the R package *MatchIt* and visualization of covariate balance was performed using the R package *cobalt.*[8,9]

After matching, all SARS-CoV-2 cases (n=30,088) and the 90,264 never-infected controls were included in the propensity-matched analysis. We repeated the PheWAS analysis using the same covariates as in the primary analysis and compared number of significant phecode associations and effects odds ratios between the full unmatched cohort (primary analysis) and the propensity-matched cohort. Results of the propensity-matched PheWAS are reported in **Table S12** and **Figure S6**.

Supplementary Tables

Table S1. Boolean logic to generate temporal-informed phenotypes

Pre-event phenotype status	Post-event phenotype status	Temporal-informed phenotype	Comment
Control	Case	Case	Phenotype cases: diagnosis is new in post-event dataset.
Control	Control	Control	Phenotype controls: patients without diagnosis code in both pre-event and post-event datasets.
Case	Any	Exclude	Exclude patients with phenotype present prior to temporal event.
Exclude	Any	Exclude	Exclude patients with phenotype exclusion prior to temporal event.
Any	Exclude	Exclude	Exclude patients with phenotype exclusion after temporal event.

Table S2. Phecode groupings to identify Charlson comorbidities

Comorbidity	Phecode
Myocardial infarction	411.2
Congestive heart failure	428, 428.1, 428.2, 428.3, 428.4
Peripheral vascular disease	443.9, 440, 440.2, 440.21, 440.22, 442.1, 442.11, 791, 459.7
Cerebrovascular disease	430, 430.1, 430.2, 430.3, 433, 433.1, 433.11, 433.12, 433.2, 433.21, 433.3, 433.31, 433.32, 433.5, 433.6, 433.8
Dementia	290, 290.1, 290.13, 290.16
Chronic pulmonary disease	497, 496.2, 496.21, 496.1, 495, 495.2, 495.1, 495.11, 496.3, 500.1, 500.2, 496, 500
Rheumatologic disease	695.42, 709.2, 709.3, 709.4, 709.5, 709.6, 709.7, 714, 714.1, 714.2, 717
Peptic ulcer disease	531, 531.1, 531.2, 531.3, 531.4, 531.5
Diabetes	250, 250.1, 250.11, 250.12, 250.13, 250.14, 250.15, 250.2, 250.21, 250.22, 250.23, 250.24, 250.25, 250.3
Mild liver disease	317.11, 070, 070.1, 070.2, 070.3, 070.4, 571.5, 571.51, 571.6
Severe liver disease	530.2, 571.8, 571.81
Hemiplegia or paraplegia	334.1, 342, 343, 344
Renal disease	401.2, 401.22, 580.1, 580.11, 580.12, 580.14, 580.3, 580.31, 580.32, 585.4, 585.34, 585.2, 588, 588.1, 588.2
Any malignancy, including lymphoma or leukemia, excluding non-melanoma skin cancer.	145, 145.1, 145.2, 145.3, 145.4, 145.5, 149, 149.1, 149.2, 149.3, 149.4, 145.9, 150, 151, 153, 153.2, 153.3, 155, 155.1, 157, 158, 159, 159.2, 159.3, 159.4, 164, 165, 165.1, 170, 170.1, 170.2, 172.11, 174, 174.1, 174.11, 174.2, 174.3, 180, 180.1, 180.3, 182, 184, 184.1, 184.2, 185, 187, 187.1, 187.2, 187.8, 189, 189.1, 189.11, 189.12, 189.2, 189.21, 189.4, 190, 191, 191.1, 191.11, 193, 194, 195, 195.1, 195.3, 200, 200.1, 201, 202, 202.2, 202.21, 202.22, 202.23, 202.24, 204, 204.1, 204.11, 204.12, 204.2, 204.21, 204.22, 204.3, 204.4, 202.22, 202.23, 202.24
Metastatic solid tumor	195.1, 198, 198.1, 198.2, 198.3, 198.4, 198.5, 198.6, 198.7
AIDS or HIV infection	071, 071.1

Table S3. Charlson comorbidities, vital signs, and laboratory data of study population

Characteristic	Never Infected	SARS-CoV-2 Positive	Overall
Number in cohort	156,017	30,088	186,105
Comorbidites prior to SARS-CoV-2 test (%) [a]			
Congestive heart failure	4,678 (3.0)	630 (2.1)	5,308 (2.9)
Diabetes	10,702 (6.9)	1,964 (6.5)	12,666 (6.8)
Myocardial infarction	2,880 (1.8)	419 (1.4)	3,299 (1.8)
Peripheral vascular disease	2,464 (1.6)	325 (1.1)	2,789 (1.5)
Cerebrovascular disease	4,289 (2.7)	610 (2.0)	4,899 (2.6)
Dementia	787 (0.5)	137 (0.5)	924 (0.5)
Chronic pulmonary disease	9,580 (6.1)	1,646 (5.5)	11,226 (6.0)
Rheumatologic disease	3,097 (2.0)	428 (1.4)	3,525 (1.9)
Peptic ulcer disease	876 (0.6)	147 (0.5)	1,023 (0.5)
Mild liver disease	4,941 (3.2)	726 (2.4)	5,667 (3.0)
Severe liver disease	1,448 (0.9)	170 (0.6)	1,618 (0.9)
Hemiplegia or paraplegia	1,035 (0.7)	153 (0.5)	1,188 (0.6)
Renal disease	3,978 (2.5)	750 (2.5)	4,728 (2.5)
Any malignancy	12,045 (7.7)	1,463 (4.9)	13,508 (7.3)
Metastatic solid tumor	2,129 (1.4)	246 (0.8)	2,375 (1.3)
AIDS or HIV infection	1,065 (0.7)	173 (0.6)	1,238 (0.7)
Vital sign or laboratory values prior to test, (mean, SD)			
Body Mass Index (kg/m^2)	24.81 (3.12)	24.81 (3.07)	24.81 (3.11)
Systolic Blood Pressure (mmHg)	121 (15)	120 (13)	121 (14)
Heart Rate (bpm)	77 (11)	77 (11)	77 (11)
Respiratory Rate (min^{-1})	17 (2)	17 (2)	17 (2)
Oxygen saturation by pulse oximetry (SpO2, %)	98 (2)	98 (2)	98 (2)
White Blood Cell Count ($10^3/\mu L$)	6.8 (1.6)	6.8 (1.6)	6.8 (1.6)
Hemoglobin (gm/dL)	13.8 (1.5)	14.0 (1.4)	13.9 (1.4)
Platelet Count ($10^3/\mu L$)	247 (55)	250 (53)	247 (54)
Serum Potassium (mEq/L)	4.1 (0.3)	4.1 (0.3)	4.1 (0.3)
Serum Creatinine (mg/dL)	0.92 (0.21)	0.92 (0.19)	0.92 (0.20)
Estimated glomerular filtration rate (ml/min)	82 (19)	83 (18)	82 (19)
Hemoglobin A1C (%)	5.6 (0.7)	5.6 (0.6)	5.6 (0.7)
Serum glucose (mg/dL)	98 (20)	98 (19)	98 (20)
Time between pre-testing and post-recovery vital sign or laboratory assessment, in days. (mean, SD).[b]			
BMI	206 (123)	214 (113)	207 (122)
Pulse	219 (124)	250 (115)	224 (123)
Blood pressure	225 (123)	249 (114)	228 (122)
Respiratory rate	244 (126)	252 (113)	246 (124)
Oxygen saturation by pulse oximetry	242 (127)	252 (112)	244 (125)
White blood cell count	268 (129)	263 (112)	267 (127)
Hemoglobin level	263 (124)	274 (113)	265 (122)
Platelet count	265 (128)	263 (112)	265 (125)
Serum potassium	258 (124)	269 (112)	260 (122)
Estimated glomerular filtration rate	259 (125)	268 (112)	261 (123)
Hemoglobin A1C	255 (125)	265 (113)	257 (123)
Serum glucose	319 (108)	314 (101)	318 (107)

[a] Phecodes used to identify Charlson comorbidities are provided in Supplementary Table S2.

[b] Pre-testing values were collected up to 180 days prior to index SARS-CoV-2 testing for all vitals and lab values, and post-recovery values were allow up to 1 year after entering the post-recovery period.

Table S4. Comparison of case retention across select phecode chapters.

Phecode chapter	Case retention (Median [IQR])	Comparator chapter	Case retention (Median [IQR])	p
neoplasms	18% [11% - 33%]	infectious diseases	40% [24% - 56%]	≤ 0.0001 *
''	''	endocrine/metabolic	28% [19% - 44%]	≤ 0.0001 *
''	''	hematopoietic	35% [24% - 45%]	0.00024 *
''	''	mental disorders	35% [24% - 56%]	≤ 0.0001 *
''	''	neurological	35% [22% - 48%]	≤ 0.0001 *
''	''	sense organs	38% [27% - 50%]	≤ 0.0001 *
''	''	circulatory system	32% [24% - 43%]	≤ 0.0001 *
''	''	respiratory	33% [26% - 51%]	≤ 0.0001 *
''	''	digestive	32% [16% - 50%]	≤ 0.0001 *
''	''	genitourinary	41% [30% - 57%]	≤ 0.0001 *
''	''	pregnancy complications	47% [37% - 60%]	≤ 0.0001 *
''	''	dermatologic	48% [28% - 59%]	≤ 0.0001 *
''	''	musculoskeletal	47% [33% - 58%]	≤ 0.0001 *
''	''	congenital anomalies	25% [18% - 43%]	0.0016
''	''	symptoms	46% [38% - 57%]	≤ 0.0001 *
''	''	injuries & poisonings	42% [32% - 55%]	≤ 0.0001 *
dermatologic	48% [28% - 59%]	infectious diseases	40% [24% - 56%]	0.32
"	"	neoplasms	18% [11% - 33%]	≤ 0.0001 *
"	"	endocrine/metabolic	28% [19% - 44%]	≤ 0.0001 *
"	"	hematopoietic	35% [24% - 45%]	0.00031 *
"	"	mental disorders	35% [24% - 56%]	0.01
"	"	neurological	35% [22% - 48%]	0.0023
"	"	sense organs	38% [27% - 50%]	0.04
"	"	circulatory system	32% [24% - 43%]	≤ 0.0001 *
"	"	respiratory	33% [26% - 51%]	0.019
"	"	digestive	32% [16% - 50%]	0.00077 *
"	"	genitourinary	41% [30% - 57%]	0.54
"	"	pregnancy complications	47% [37% - 60%]	0.56
"	"	musculoskeletal	47% [33% - 58%]	0.89
"	"	congenital anomalies	25% [18% - 43%]	0.0012
"	"	symptoms	46% [38% - 57%]	0.59
"	"	injuries & poisonings	42% [32% - 55%]	0.22
musculoskeletal	47% [33% - 58%]	infectious diseases	40% [24% - 56%]	0.26
"	"	neoplasms	18% [11% - 33%]	≤ 0.0001 *
"	"	endocrine/metabolic	28% [19% - 44%]	≤ 0.0001 *
"	"	hematopoietic	35% [24% - 45%]	≤ 0.0001 *
"	"	mental disorders	35% [24% - 56%]	0.0049
"	"	neurological	35% [22% - 48%]	0.00083 *
"	"	sense organs	38% [27% - 50%]	0.013
"	"	circulatory system	32% [24% - 43%]	≤ 0.0001 *
"	"	respiratory	33% [26% - 51%]	0.006
"	"	digestive	32% [16% - 50%]	0.0002 *
"	"	genitourinary	41% [30% - 57%]	0.33
"	"	pregnancy complications	47% [37% - 60%]	0.48
"	"	dermatologic	48% [28% - 59%]	0.89
"	"	congenital anomalies	25% [18% - 43%]	0.00063 *
"	"	symptoms	46% [38% - 57%]	0.68
"	"	injuries & poisonings	42% [32% - 55%]	0.13

congenital anomalies	25% [18% - 43%]	infectious diseases	40% [24% - 56%]	0.057
"	"	neoplasms	18% [11% - 33%]	0.0016
"	"	endocrine/metabolic	28% [19% - 44%]	0.82
"	"	hematopoietic	35% [24% - 45%]	0.98
"	"	mental disorders	35% [24% - 56%]	0.29
"	"	neurological	35% [22% - 48%]	0.41
"	"	sense organs	38% [27% - 50%]	0.048
"	"	circulatory system	32% [24% - 43%]	0.6
"	"	respiratory	33% [26% - 51%]	0.12
"	"	digestive	32% [16% - 50%]	0.48
"	"	genitourinary	41% [30% - 57%]	0.003
"	"	pregnancy complications	47% [37% - 60%]	0.043
"	"	dermatologic	48% [28% - 59%]	0.0012
"	"	musculoskeletal	47% [33% - 58%]	0.00063 *
"	"	symptoms	46% [38% - 57%]	0.0018
"	"	injuries & poisonings	42% [32% - 55%]	0.038
symptoms	46% [38% - 57%]	infectious diseases	40% [24% - 56%]	0.16
"	"	neoplasms	18% [11% - 33%]	≤ 0.0001 *
"	"	endocrine/metabolic	28% [19% - 44%]	≤ 0.0001 *
"	"	hematopoietic	35% [24% - 45%]	0.00015 *
"	"	mental disorders	35% [24% - 56%]	0.0052
"	"	neurological	35% [22% - 48%]	0.0016
"	"	sense organs	38% [27% - 50%]	0.015
"	"	circulatory system	32% [24% - 43%]	≤ 0.0001 *
"	"	respiratory	33% [26% - 51%]	0.0049
"	"	digestive	32% [16% - 50%]	0.001
"	"	genitourinary	41% [30% - 57%]	0.21
"	"	pregnancy complications	47% [37% - 60%]	0.34
"	"	dermatologic	48% [28% - 59%]	0.59
"	"	musculoskeletal	47% [33% - 58%]	0.68
"	"	congenital anomalies	25% [18% - 43%]	0.0018
"	"	injuries & poisonings	42% [32% - 55%]	0.1

* $p < 0.001$

Table S5. PheWAS summary under naive post-acute phenotyping: top 100 associations.

Phecode	Description	Odds Ratio (95% CI)	p value	No. cases	No. controls
585.1	Acute renal failure	5.72 (4.91-6.68)	5.07E-109	1,054	183,082
427.5	Arrhythmia (cardiac) NOS	6.85 (5.67-8.26)	1.92E-89	632	178,215
292.4	Altered mental status	12.5 (9.55-16.3)	8.91E-77	258	187,128
418.1	Precordial pain	4.10 (3.53-4.77)	8.73E-75	1,034	183,404
285.1	Acute posthemorrhagic anemia	16.3 (12.0-22.0)	5.43E-73	189	183,114
458.9	Hypotension NOS	10.3 (7.96-13.4)	5.54E-70	274	189,734
401.1	Essential hypertension	1.63 (1.54-1.72)	5.60E-68	15,711	163,382
401.22	Hypertensive chronic kidney disease	6.19 (5.03-7.62)	1.16E-66	738	163,382
285	Other anemias	3.37 (2.93-3.87)	6.74E-65	1,443	183,114
509.1	Respiratory failure	5.07 (4.17-6.16)	7.60E-60	600	188,161
994.2	Sepsis	11.5 (8.50-15.7)	2.06E-55	196	191,158
1013	Asphyxia and hypoxemia	9.14 (6.93-12.1)	2.78E-55	240	190,957
348.8	Encephalopathy, not elsewhere classified	14.6 (10.4-20.4)	5.97E-55	157	184,955
38	Septicemia	12.6 (9.16-17.4)	7.00E-54	176	188,677
772.3	Muscle weakness	5.01 (4.08-6.15)	9.28E-54	536	188,033
272.1	Hyperlipidemia	2.29 (2.06-2.55)	7.06E-53	3,137	171,670
276.6	Fluid overload	8.37 (6.29-11.1)	2.65E-48	245	186,845
512.9	Other dyspnea	2.47 (2.18-2.79)	1.44E-46	1,835	171,917
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	5.25 (4.18-6.59)	2.00E-46	368	108,651
276.41	Acidosis	4.87 (3.91-6.05)	6.44E-46	513	186,845
338.1	Acute pain	2.64 (2.30-3.02)	1.07E-43	1,652	177,888
512.7	Shortness of breath	1.95 (1.78-2.15)	1.43E-42	3,451	171,917
250.22	Type 2 diabetes with renal manifestations	3.47 (2.90-4.15)	8.98E-42	1,181	177,510
38.3	Bacteremia	15.1 (10.2-22.4)	1.11E-41	118	188,677
276.5	Hypovolemia	6.93 (5.23-9.17)	1.20E-41	261	186,845
276.12	Hyposmolality and/or hyponatremia	4.89 (3.88-6.17)	3.26E-41	433	186,845
278.11	Morbid obesity	2.10 (1.88-2.35)	2.05E-39	2,335	182,589
260.2	Severe protein-calorie malnutrition	7.91 (5.81-10.8)	2.68E-39	212	183,599
530.11	GERD	1.89 (1.72-2.08)	6.32E-39	3,958	178,019
418	Nonspecific chest pain	2.09 (1.87-2.35)	1.32E-37	2,264	183,404
250.2	Type 2 diabetes	1.85 (1.68-2.04)	1.78E-36	5,316	177,510
401.2	Hypertensive heart and/or renal disease	6.69 (4.98-9.01)	3.48E-36	373	163,382
585.3	Chronic renal failure [CKD]	3.57 (2.91-4.38)	1.94E-34	787	183,082
276.13	Hyperpotassemia	3.56 (2.90-4.38)	1.63E-33	621	186,845
1010.6	Reproductive and maternal health services	2.08 (1.84-2.35)	2.25E-32	1,598	189,380
285.21	Anemia in chronic kidney disease	3.64 (2.94-4.52)	3.35E-32	629	183,114
428.1	Congestive heart failure (CHF) NOS	4.48 (3.49-5.75)	6.83E-32	427	184,871
401.21	Hypertensive heart disease	3.63 (2.92-4.52)	3.78E-31	701	163,382
288.2	Elevated white blood cell count	4.41 (3.43-5.67)	8.23E-31	344	187,148
509.8	Dependence on respirator [Ventilator] or supplemental oxygen	7.42 (5.24-10.5)	1.37E-29	160	188,161
665	Obstetrical/birth trauma	7.96 (5.54-11.4)	2.50E-29	152	191,347
411.2	Myocardial infarction	3.44 (2.75-4.31)	5.97E-27	607	182,562
327.32	Obstructive sleep apnea	1.75 (1.58-1.94)	7.65E-27	3,401	179,431
508	Pulmonary collapse; interstitial and compensatory emphysema	6.58 (4.66-9.29)	8.33E-27	176	188,161
426	Cardiac conduction disorders	9.15 (5.95-14.1)	6.94E-24	106	178,215
591	Urinary tract infection	2.29 (1.95-2.69)	1.75E-23	1,145	184,472
798	Malaise and fatigue	1.63 (1.48-1.80)	2.00E-23	3,445	178,495
276.14	Hypopotassemia	3.36 (2.65-4.27)	2.91E-23	466	186,845
260.3	Adult failure to thrive	8.99 (5.81-13.9)	5.90E-23	101	183,599
789	Nausea and vomiting	1.83 (1.62-2.07)	1.44E-22	2,139	184,375
994.21	Septic shock	18.0 (10.1-32.4)	2.70E-22	53	191,158

797	Shock	16.2 (9.16-28.5)	7.79E-22	55	191,476
297.1	Suicidal ideation	5.50 (3.89-7.80)	7.90E-22	192	170,204
655	Known or suspected fetal abnormality affecting management of mother	3.41 (2.65-4.38)	1.15E-21	310	108,936
272.11	Hypercholesterolemia	1.82 (1.61-2.06)	1.24E-21	2,281	171,670
244.4	Hypothyroidism NOS	1.69 (1.51-1.88)	4.88E-21	2,802	182,090
250.24	Type 2 diabetes with neurological manifestations	2.29 (1.93-2.73)	5.53E-21	1,344	177,510
1010	Other tests	3.50 (2.69-4.56)	1.08E-20	308	189,547
427.7	Tachycardia NOS	2.55 (2.10-3.11)	1.41E-20	663	178,215
646	Other complications of pregnancy NEC	5.10 (3.60-7.22)	3.68E-20	160	109,219
278.1	Obesity	1.90 (1.66-2.18)	3.98E-20	1,549	182,589
644	Anemia during pregnancy	10.0 (6.09-16.6)	1.62E-19	73	109,431
274.1	Gout	3.25 (2.51-4.21)	3.73E-19	351	189,924
654.1	Abnormality of organs and soft tissues of pelvis complicating pregnancy, childbirth, or the puerperium	4.04 (2.97-5.50)	5.69E-19	209	108,966
136	Other infectious and parasitic diseases	6.74 (4.41-10.3)	1.17E-18	104	191,038
428.4	Heart failure with preserved EF [Diastolic heart failure]	2.39 (1.96-2.91)	7.53E-18	1,167	184,871
663	Umbilical cord complications during labor and delivery	21.8 (10.8-44.1)	1.02E-17	44	191,521
638	Other high-risk pregnancy	2.38 (1.95-2.90)	1.06E-17	584	190,705
290.2	Delirium due to conditions classified elsewhere	20.4 (10.2-40.6)	1.07E-17	37	187,128
411.4	Coronary atherosclerosis	1.60 (1.44-1.79)	2.26E-17	4,362	182,562
585.32	End stage renal disease	2.69 (2.14-3.38)	2.37E-17	722	183,082
588.1	Renal osteodystrophy	3.12 (2.39-4.06)	3.32E-17	456	183,082
427.21	Atrial fibrillation	1.71 (1.51-1.94)	5.90E-17	3,184	178,215
284.1	Pancytopenia	3.94 (2.85-5.45)	1.01E-16	256	183,114
785	Abdominal pain	1.42 (1.31-1.54)	1.27E-16	5,027	180,428
452.2	Deep vein thrombosis [DVT]	2.57 (2.05-3.21)	1.63E-16	569	185,983
480	Pneumonia	2.83 (2.21-3.63)	1.81E-16	387	188,133
653	Problems associated with amniotic cavity and membranes	13.4 (7.22-24.8)	1.86E-16	48	108,966
585.34	Chronic Kidney Disease, Stage IV	2.91 (2.26-3.76)	2.25E-16	537	183,082
590	Pyelonephritis	5.81 (3.81-8.86)	2.95E-16	113	184,472
851	Complications of transplants and reattached limbs	2.38 (1.94-2.94)	3.18E-16	788	187,602
287.3	Thrombocytopenia	2.99 (2.30-3.90)	5.75E-16	382	188,946
507	Pleurisy; pleural effusion	3.24 (2.43-4.31)	7.34E-16	351	188,161
280.1	Iron deficiency anemias, unspecified or not due to blood loss	1.91 (1.63-2.24)	1.36E-15	1,330	183,114
578.9	Hemorrhage of gastrointestinal tract	5.45 (3.59-8.29)	2.10E-15	123	188,746
260	Protein-calorie malnutrition	3.39 (2.50-4.58)	2.90E-15	302	183,599
41	Bacterial infection NOS	3.15 (2.36-4.21)	5.79E-15	288	188,677
359.2	Myopathy	8.12 (4.79-13.8)	7.43E-15	71	188,619
285.2	Anemia of chronic disease	6.90 (4.24-11.2)	8.32E-15	89	183,114
250.42	Other abnormal glucose	2.16 (1.78-2.63)	8.87E-15	805	177,510
272.13	Mixed hyperlipidemia	1.44 (1.31-1.58)	8.95E-15	5,236	171,670
38.2	Gram positive septicemia	16.5 (8.08-33.7)	1.43E-14	36	188,677
41.4	E. coli	13.3 (6.86-25.9)	2.25E-14	41	188,677
41.9	Infection with drug-resistant microorganisms	15.1 (7.53-30.3)	2.26E-14	37	188,677
1090	Acquired absence of organs	2.55 (2.00-3.24)	2.80E-14	569	189,321
642	Hypertension complicating pregnancy, childbirth, and the puerperium	5.51 (3.53-8.61)	6.49E-14	93	109,399
567	Peritonitis and retroperitoneal infections	4.84 (3.20-7.32)	8.48E-14	135	187,178
276.11	Hyperosmolality and/or hypernatremia	34.2 (13.4-87.6)	1.65E-13	22	186,845
532	Dysphagia	1.84 (1.56-2.16)	4.40E-13	1,445	178,019

Table S6. Temporal-informed PheWAS by race/ethnicity subgroups.

Phecode[a]	Description	Odds Ratio	95% CI	p value	No. cases	No. controls
White, Non-Hispanic patients (N=125,938)						
512.9	Other dyspnea	3.22	(2.59-4.00)	7.73E-26	627	59,636
512.7	Shortness of breath	2.81	(2.29-3.44)	1.66E-23	760	59,636
359.2	Myopathy	29.9	(12.2-73.3)	1.26E-13	25	113,531
427.9	Palpitations	2.35	(1.85-2.97)	1.26E-12	468	84,577
569.2	Gastrointestinal complications of surgery	5.69	(3.48-9.30)	3.93E-12	85	107,583
418.1	Precordial pain	3.78	(2.59-5.52)	5.31E-12	194	87,353
278.11	Morbid obesity	2.32	(1.82-2.96)	1.08E-11	437	100,068
136	Other infectious and parasitic diseases	11.4	(5.65-23.0)	1.14E-11	38	119,882
509.1	Respiratory failure	6.32	(3.66-10.9)	3.92E-11	86	100,780
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	3.77	(2.53-5.62)	6.23E-11	124	63,357
427.21	Atrial fibrillation	2.67	(1.95-3.64)	6.86E-10	382	84,577
585.1	Acute renal failure	3.25	(2.18-4.86)	8.87E-09	218	100,977
327.32	Obstructive sleep apnea	2.05	(1.60-2.62)	1.14E-08	526	94,165
418	Nonspecific chest pain	1.97	(1.55-2.50)	2.52E-08	530	87,353
782.3	Edema	2.27	(1.67-3.08)	1.78E-07	331	108,722
1010	Other tests	3.26	(2.09-5.11)	2.23E-07	111	110,956
250.2	Type 2 diabetes	2.13	(1.60-2.84)	2.41E-07	406	93,983
599.2	Retention of urine	3.39	(2.13-5.41)	2.95E-07	148	94,495
1013	Asphyxia and hypoxemia	6.28	(3.11-12.7)	3.09E-07	43	114,386
350.1	Abnormal involuntary movements	2.55	(1.76-3.69)	6.66E-07	205	111,360
249	Secondary diabetes mellitus	8.71	(3.62-21.0)	1.39E-06	44	93,983
452.2	Deep vein thrombosis [DVT]	3.45	(2.08-5.74)	1.76E-06	105	104,177
452	Other venous embolism and thrombosis	4.49	(2.41-8.37)	2.22E-06	68	104,177
420.1	Myocarditis	11.4	(4.05-31.9)	3.84E-06	17	115,711
292	Neurological disorders	2.47	(1.68-3.62)	3.90E-06	188	104,296
514	Abnormal findings examination of lungs	2.42	(1.66-3.53)	4.80E-06	284	105,200
646	Other complications of pregnancy NEC	4.09	(2.22-7.51)	5.71E-06	50	65,813
1010.6	Reproductive and maternal health services	1.71	(1.35-2.17)	8.43E-06	408	114,589
284.1	Pancytopenia	3.91	(2.11-7.22)	1.38E-05	73	94,032
727.1	Synovitis and tenosynovitis	2.31	(1.58-3.37)	1.47E-05	212	92,564
386.9	Dizziness and giddiness	1.74	(1.35-2.25)	1.69E-05	489	103,688
781	Symptoms involving nervous and musculoskeletal systems	3.01	(1.82-4.98)	1.83E-05	117	117,833
433.1	Occlusion and stenosis of precerebral arteries	3.66	(1.99-6.73)	3.16E-05	83	111,257
401.1	Essential hypertension	1.45	(1.22-1.73)	3.38E-05	1,268	74,057
285.2	Anemia of chronic disease	11.0	(3.53-34.2)	3.54E-05	17	94,032
649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	4.61	(2.21-9.61)	4.58E-05	36	63,357
Black, Non-Hispanic patients (N=19,936)						
798.1	Chronic fatigue syndrome	7.63	(3.39-17.2)	8.98E-07	29	12,789
569.2	Gastrointestinal complications of surgery	8.13	(3.50-18.9)	1.07E-06	24	16,681
278.11	Morbid obesity	2.95	(1.87-4.65)	3.43E-06	106	13,894
638	Other high-risk pregnancy	4.38	(2.19-8.78)	3.07E-05	37	18,235
250.42	Other abnormal glucose	4.76	(2.23-10.1)	5.51E-05	42	13,745
280.2	Iron deficiency anemia secondary to blood loss	5.08	(2.20-11.7)	1.39E-04	28	12,275

[a] A list of ICD-10-CM codes included in each phecode is available at: https://phewascatalog.org/phecodes_icd10cm [10]

Table S7. Temporal-informed PheWAS by sex.

Temporal-informed PheWAS among both male and female patients.											
		Females					Males				
Phenotype	Description	Odds Ratio	95% CI	p value	No. cases	No. controls	Odds Ratio	95% CI	p value	No. cases	No. controls
512.9	Other dyspnea	2.75	(2.13-3.55)	9.38×10^{-15}	456	52,865	3.46	(2.61-4.58)	6.56×10^{-18}	355	41,071
512.7	Shortness of breath	2.24	(1.79-2.82)	3.35×10^{-12}	597	52,865	2.94	(2.22-3.88)	3.81×10^{-14}	391	41,071
278.11	Morbid obesity	2.29	(1.82-2.87)	1.06×10^{-12}	464	85,699	2.50	(1.7-3.69)	3.76×10^{-06}	160	69,162
359.2	Myopathy	25.2	(7.23-88.2)	4.18×10^{-07}	15	100,096	18.0	(6.24-51.9)	8.97×10^{-08}	18	74,767
418.1	Precordial pain	2.99	(2-4.47)	8.81×10^{-08}	174	79,908	3.66	(2.2-6.07)	5.16×10^{-07}	104	58,629
418	Nonspecific chest pain	1.93	(1.51-2.47)	1.46×10^{-07}	444	79,908	2.13	(1.56-2.9)	1.84×10^{-06}	302	58,629
585.1	Acute renal failure	3.61	(2.27-5.76)	6.15×10^{-08}	151	92,868	2.76	(1.72-4.42)	2.47×10^{-05}	158	64,607

Temporal-informed PheWAS among female patients only.											
		Females					Males				
Phenotype	Description	Odds Ratio	95% CI	p value	No. cases	No. controls	Odds Ratio	95% CI	p value	No. cases	No. controls
569.2	Gastrointestinal complications	6.54	(4.24-10.1)	2.51×10^{-17}	94	95,244	5.78	(1.96-17)	0.0014	22	71,581
136	Other infectious and parasitic diseases	16.0	(7.57-33.7)	3.55×10^{-13}	35	103,699	3.33	(1.19-9.26)	0.0214	19	78,267
427.9	Palpitations	2.35	(1.86-2.97)	1.07×10^{-12}	441	80,047	1.67	(1.12-2.47)	0.011	187	57,039
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	2.94	(2.15-4.03)	1.71×10^{-11}	169	95,518	-	-	-	-	-
646	Other complications of pregnancy NEC	4.27	(2.66-6.86)	2.06×10^{-09}	69	99,542	-	-	-	-	-
1010.6	Reproductive and maternal health services	1.75	(1.44-2.12)	9.99×10^{-09}	591	92,952	-	-	-	-	-
638	Other high-risk pregnancy	2.19	(1.67-2.86)	1.34×10^{-08}	312	98,919	-	-	-	-	-
1010	Other tests	3.19	(2.09-4.86)	7.43×10^{-08}	116	92,964	2.7	(1.22-5.95)	0.014	39	76,383
420.1	Myocarditis	16.1	(5.52-47)	3.67×10^{-07}	15	102,140	-	-	-	5	74,863
401.1	Essential hypertension	1.63	(1.34-1.98)	1.01×10^{-06}	865	74,906	1.22	(0.98-1.52)	0.0725	833	48,001
644	Anemia during pregnancy	4.72	(2.5-8.92)	1.81×10^{-06}	38	101,761	-	-	-	-	-
452.2	Deep vein thrombosis [DVT]	4.37	(2.38-8.03)	1.96×10^{-06}	66	93,371	2.37	(1.27-4.44)	0.007	72	69,340
285	Other anemias	2.34	(1.64-3.32)	2.21×10^{-06}	259	82,899	1.69	(1.11-2.58)	0.0149	214	63,606
649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	3.52	(2.08-5.98)	3.11×10^{-06}	57	95,518	-	-	-	-	-
292	Neurological disorders	2.70	(1.76-4.14)	5.11×10^{-06}	135	93,719	2.01	(1.2-3.37)	0.00826	107	68,515

671	Venous/cerebrovascular complications embolism in pregnancy and the puerperium	9.77	(3.61-26.4)	7.09×10^{-06}	17	103,586	-	-	-	-	-
782.3	Edema	2.25	(1.57-3.22)	9.51×10^{-06}	231	95,466	1.91	(1.27-2.88)	0.00182	193	72,718
348.8	Encephalopathy, not elsewhere classified	12.1	(3.95-37.1)	1.27×10^{-05}	17	92,273	3.40	(0.91-12.7)	0.0694	15	68,246
284.1	Pancytopenia	5.01	(2.42-10.4)	1.45×10^{-05}	46	82,899	2.02	(0.84-4.84)	0.114	48	63,606

Temporal-informed PheWAS among male patients only.											
		Females					**Males**				
Phenotype	**Description**	**Odds Ratio**	**95% CI**	**p value**	**No. cases**	**No. controls**	**Odds Ratio**	**95% CI**	**p value**	**No. cases**	**No. controls**
509.1	Respiratory failure	3.23	(1.45-7.17)	0.00401	50	93,041	12.4	(6.53-23.5)	1.27×10^{-14}	51	64,751
427.21	Atrial fibrillation	1.96	(1.23-3.13)	0.00462	170	80,047	3.27	(2.28-4.68)	9.13×10^{-11}	273	57,039
840	Sprains and strains	1.10	(0.80-1.51)	0.57	348	91,895	2.28	(1.64-3.18)	1.08×10^{-06}	249	69,749
514	Abnormal findings examination of lungs	1.76	(1.09-2.82)	0.02	183	95,645	3.12	(1.95-5)	2.04×10^{-06}	167	67,924
278.1	Obesity	1.41	(1.07-1.85)	0.015	401	85,699	2.41	(1.65-3.51)	5.73×10^{-06}	165	69,162
350.1	Abnormal involuntary movements	2.22	(1.44-3.43)	0.0003	149	96,901	2.99	(1.85-4.84)	8.5×10^{-06}	107	73,586
1013	Asphyxia and hypoxemia	4.40	(1.73-11.2)	0.002	27	101,438	7.13	(2.89-17.6)	1.98×10^{-05}	25	74,001
295.1	Schizophrenia	-	-	-	9	66,054	6.71	(2.67-16.9)	5.13×10^{-05}	26	58,256

Table S8. Temporal-informed PheWAS by onset of diagnosis.

Phenotype	Description	Early-presenting post-acute phenotypes [a]					Late-presenting post-acute phenotypes [b]				
		Odds Ratio	95% CI	p value	No. cases	No. controls	Odds Ratio	95% CI	p value	No. cases	No. controls
512.9	Other dyspnea	3.89	(2.80-5.41)	7.14E-16	227	96,877	2.79	(2.22-3.51)	1.13E-18	584	90,181
569.2	Gastrointestinal complications	9.81	(5.21-18.5)	1.62E-12	44	167,795	5.11	(3.06-8.53)	4.31E-10	72	160,967
509.1	Respiratory failure	12.8	(6.24-26.2)	3.25E-12	43	158,293	4.42	(2.22-8.82)	2.38E-05	58	152,379
427.9	Palpitations	3.02	(2.11-4.32)	1.68E-09	172	138,826	1.86	(1.46-2.37)	6.56E-07	456	132,806
512.7	Shortness of breath	2.52	(1.86-3.42)	2.83E-09	292	96,877	2.52	(2.03-3.11)	2.05E-17	696	90,181
644	Anemia during pregnancy	19.0	(6.26-57.4)	1.97E-07	16	101,815	-	-	-	-	-
514	Abnormal findings examination of lungs	3.47	(2.06-5.85)	2.84E-06	113	164,532	-	-	-	-	-
285	Other anemias	2.83	(1.83-4.39)	3.41E-06	157	147,698	-	-	-	-	-
418.1	Precordial pain	4.42	(2.34-8.37)	4.74E-06	56	140,112	2.94	(2.05-4.21)	4.7E-09	222	134,057
136	Other infectious and parasitic diseases	11.9	(3.81-37.1)	2.03E-05	15	182,227	8.46	(4.26-16.8)	1.03E-09	39	175,181
585.1	Acute renal failure	3.24	(1.84-5.69)	4.69E-05	107	158,429	3.26	(2.19-4.85)	5.22E-09	202	152,145
359.2	Myopathy	12.7	(3.74-43.5)	4.78E-05	16	175,579	28.1	(9.94-79.6)	3.23E-10	17	168,524
452	Other venous embolism and thrombosis	5.49	(2.36-12.7)	7.37E-05	32	164,545	-	-	-	-	-
401.1	Essential hypertension	1.62	(1.27-2.06)	9.52E-05	617	124,813	-	-	-	-	-
278.11	Morbid obesity	-	-	-	-	-	2.49	(1.98-3.14)	9.64E-15	443	149,748
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	-	-	-	-	-	3.66	(2.54-5.27)	3.26E-12	145	92,130
427.21	Atrial fibrillation	-	-	-	-	-	3.23	(2.30-4.56)	1.88E-11	263	132,806
418	Nonspecific chest pain	-	-	-	-	-	1.99	(1.59-2.49)	1.41E-09	566	134,057
292	Neurological disorders	-	-	-	-	-	2.97	(2.09-4.22)	1.51E-09	192	156,592
646	Other complications of pregnancy NEC	-	-	-	-	-	5.06	(2.86-8.93)	2.3E-08	55	95,983
1010	Other tests	-	-	-	-	-	3.23	(2.11-4.94)	7.01E-08	121	163,187
638	Other high-risk pregnancy	-	-	-	-	-	2.23	(1.66-2.98)	8.46E-08	269	172,113
395.6	Heart valve replaced	-	-	-	-	-	11.0	(4.43-27.5)	2.53E-07	23	160,821
1010.6	Reproductive and maternal health services	-	-	-	-	-	1.74	(1.41-2.15)	2.55E-07	496	166,455
350.1	Abnormal involuntary movements	-	-	-	-	-	2.61	(1.81-3.77)	3.19E-07	185	164,415
782.3	Edema	-	-	-	-	-	2.22	(1.62-3.03)	5.54E-07	301	162,355
781	Symptoms involving nervous and musculoskeletal systems	-	-	-	-	-	3.44	(2.11-5.60)	7.36E-07	105	173,467
433.1	Occlusion and stenosis of precerebral arteries	-	-	-	-	-	4.72	(2.47-9.00)	2.6E-06	64	165,557
260	Protein-calorie malnutrition	-	-	-	-	-	4.23	(2.29-7.82)	4.06E-06	69	151,935
278.1	Obesity	-	-	-	-	-	1.80	(1.40-2.32)	4.28E-06	437	149,748
348.8	Encephalopathy, not elsewhere classified	-	-	-	-	-	11.6	(4.06-33.0)	4.6E-06	19	154,973

649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	-	-	-	-	-	4.91	(2.48-9.72)	5.01E-06	39	92,130
79	Viral infection	-	-	-	-	-	3.97	(2.18-7.23)	6.42E-06	74	144,377
244.4	Hypothyroidism NOS	-	-	-	-	-	2.10	(1.51-2.92)	1.04E-05	261	154,957
327.32	Obstructive sleep apnea	-	-	-	-	-	1.77	(1.37-2.29)	1.39E-05	479	145,658
276.13	Hyperpotassemia	-	-	-	-	-	3.18	(1.87-5.41)	1.96E-05	108	144,713
199	Neoplasm of uncertain behavior	-	-	-	-	-	3.23	(1.87-5.60)	2.83E-05	84	166,548
428.4	Heart failure with preserved EF	-	-	-	-	-	2.78	(1.69-4.56)	5.17E-05	142	159,733
285.22	Anemia in neoplastic disease	-	-	-	-	-	2.88	(1.73-4.82)	5.18E-05	117	141,744
701.2	Scar conditions and fibrosis of skin	-	-	-	-	-	3.04	(1.77-5.23)	5.44E-05	102	166,637

[a] First post-acute diagnosis made within 60 days of recovery following COVID-19 testing
[b] First post-acute diagnosis made after 60 days of recovery following COVID-19 testing

Table S9. Sensitivity analysis: adults with a least one follow-up visit in post-acute period.

Phecode	Description	Odds Ratio	95% CI	p value	No. cases	No. controls
512.9	Other dyspnea	3.05	(2.53-3.68)	3.18E-31	811	50,602
512.7	Shortness of breath	2.52	(2.11-3.00)	6.45E-25	988	50,602
569.2	Gastrointestinal complications	5.86	(3.94-8.73)	3.00E-18	116	97,430
278.11	Morbid obesity	2.24	(1.84-2.72)	6.34E-16	623	88,502
509.1	Respiratory failure	6.99	(4.28-11.4)	6.96E-15	101	93,258
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	3.65	(2.61-5.10)	3.49E-14	167	60,558
359.2	Myopathy	19.1	(8.69-41.9)	2.03E-13	33	103,726
136	Other infectious and parasitic diseases	8.63	(4.84-15.4)	2.73E-13	54	110,018
418.1	Precordial pain	3.03	(2.22-4.13)	2.7E-12	278	78,592
427.9	Palpitations	2.03	(1.66-2.48)	5.26E-12	627	76,834
418	Nonspecific chest pain	1.95	(1.61-2.36)	8.73E-12	745	78,592
646	Other complications of pregnancy NEC	5.35	(3.23-8.87)	8.21E-11	69	63,503
427.21	Atrial fibrillation	2.53	(1.91-3.36)	1.29E-10	443	76,834
585.1	Acute renal failure	2.93	(2.11-4.08)	1.59E-10	309	91,680
1010.6	Reproductive and maternal health services	1.79	(1.47-2.19)	6.06E-09	549	104,034
1010	Other tests	2.97	(2.05-4.29)	8.00E-09	155	99,632
644	Anemia during pregnancy	6.69	(3.37-13.3)	5.62E-08	38	65,119
350.1	Abnormal involuntary movements	2.40	(1.74-3.30)	7.54E-08	256	101,168
671	Venous/cerebrovascular complications embolism in pregnancy and the puerperium	19.2	(6.51-56.6)	8.6E-08	17	66,122
638	Other high-risk pregnancy	2.11	(1.61-2.78)	9.57E-08	298	107,781
452.2	Deep vein thrombosis [DVT]	3.11	(2.02-4.79)	2.61E-07	138	93,807
649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	4.40	(2.50-7.75)	2.68E-07	57	60,558
292	Neurological disorders	2.33	(1.68-3.22)	4.11E-07	242	96,775
1013	Asphyxia and hypoxemia	5.24	(2.76-9.95)	4.28E-07	52	105,514
782.3	Edema	1.99	(1.52-2.60)	4.72E-07	424	98,285
781	Symptoms involving nervous and musculoskeletal systems	2.89	(1.90-4.39)	7.16E-07	151	108,474
599.2	Retention of urine	2.72	(1.82-4.08)	1.21E-06	184	82,558
285	Other anemias	1.92	(1.47-2.51)	1.87E-06	473	83,038
401.1	Essential hypertension	1.42	(1.23-1.64)	2.71E-06	1697	66,158
587	Kidney replaced by transpant	26.5	(6.61-106.)	3.71E-06	22	91,680
514	Abnormal findings examination of lungs	2.18	(1.57-3.04)	3.75E-06	350	96,271
420.1	Myocarditis	9.03	(3.47-23.5)	6.64E-06	20	105,934
250.2	Type 2 diabetes	1.73	(1.36-2.20)	9.29E-06	572	83,868
278.1	Obesity	1.64	(1.31-2.04)	1.07E-05	565	88,502
327.32	Obstructive sleep apnea	1.62	(1.30-2.01)	1.54E-05	669	84,158
348.8	Encephalopathy, not elsewhere classified	5.96	(2.64-13.5)	1.79E-05	32	94,836
502	Postinflammatory pulmonary fibrosis	5.34	(2.44-11.7)	2.79E-05	40	93,258
653	Problems associated with amniotic cavity and membranes	7.02	(2.75-17.9)	4.49E-05	19	61,980
655	Known or suspected fetal abnormality affecting management of mother	2.13	(1.48-3.06)	4.55E-05	158	59,807

Table S10. Sensitivity analysis: phenotype case definition of one post-acute phecode.

Phecode	Description	Odds Ratio	95% CI	p value	No. cases	No. controls
512.9	Other dyspnea	2.13	(1.89-2.41)	8.88E-35	2202	93,936
136	Other infectious and parasitic diseases	6.51	(4.74-8.92)	3.24E-31	185	181,966
418.1	Precordial pain	2.62	(2.19-3.14)	6.66E-26	978	138,537
509.1	Respiratory failure	4.85	(3.61-6.50)	8.38E-26	304	157,792
278.11	Morbid obesity	1.81	(1.59-2.05)	3.45E-20	1678	154,861
646	Other complications of pregnancy NEC	3.61	(2.73-4.78)	3.16E-19	240	99,542
512.7	Shortness of breath	1.60	(1.44-1.77)	5.86E-19	3182	93,936
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	2.97	(2.32-3.79)	3.25E-18	323	95,518
647	Infectious and parasitic complications affecting pregnancy	12.33	(6.84-22.2)	7.03E-17	65	102,065
636	Early or threatened labor; hemorrhage in early pregnancy	7.35	(4.59-11.8)	8.94E-17	79	95,926
1010	Other tests	1.77	(1.55-2.03)	1.92E-16	1365	169,347
569.2	Gastrointestinal complications	3.83	(2.77-5.29)	3.59E-16	221	166,825
285	Other anemias	1.88	(1.61-2.20)	7.55E-16	1515	146,505
585.1	Acute renal failure	2.47	(1.98-3.08)	1.16E-15	792	157,475
427.21	Atrial fibrillation	2.37	(1.90-2.96)	1.52E-14	793	137,086
348.8	Encephalopathy, not elsewhere classified	5.32	(3.40-8.34)	2.85E-13	113	160,519
359.2	Myopathy	5.81	(3.62-9.32)	3.38E-13	91	174,863
782.3	Edema	1.76	(1.51-2.04)	3.47E-13	1396	168,184
644	Anemia during pregnancy	4.62	(3.00-7.10)	3.19E-12	111	101,761
327.32	Obstructive sleep apnea	1.63	(1.42-1.87)	3.47E-12	1737	150,608
798.1	Chronic fatigue syndrome	1.81	(1.53-2.15)	4.63E-12	946	131,770
418	Nonspecific chest pain	1.49	(1.33-1.67)	4.65E-12	2233	138,537
427.5	Arrhythmia (cardiac) NOS	2.13	(1.72-2.64)	5.6E-12	870	137,086
427.9	Palpitations	1.65	(1.43-1.92)	1.45E-11	1306	137,086
278.1	Obesity	1.56	(1.37-1.77)	1.49E-11	1802	154,861
452.2	Deep vein thrombosis [DVT]	2.7	(2.01-3.62)	4.58E-11	327	162,711
292	Neurological disorders	1.97	(1.61-2.41)	4.75E-11	668	162,234
671	Venous/cerebrovascular complications embolism in pregnancy and the puerperium	8.74	(4.54-16.8)	9.23E-11	45	103,586
772.3	Muscle weakness	1.79	(1.49-2.14)	2.86E-10	1139	155,495
661	Fetal distress and abnormal forces of labor	9.70	(4.70-20.0)	7.75E-10	34	181,989
401.1	Essential hypertension	1.36	(1.23-1.49)	8.65E-10	4049	122,907
286.7	Other and unspecified coagulation defects	2.76	(1.98-3.86)	2.63E-09	283	168,886
514	Abnormal findings examination of lungs	1.65	(1.40-1.94)	2.84E-09	1442	163,569
1013	Asphyxia and hypoxemia	2.90	(2.04-4.12)	3.2E-09	215	175,439
655	Known or suspected fetal abnormality affecting management of mother	1.94	(1.55-2.43)	8.5E-09	433	94,447
509.2	Respiratory insufficiency	8.21	(3.97-17.0)	1.27E-08	40	157,792
260	Protein-calorie malnutrition	2.26	(1.70-3.02)	2.48E-08	438	157,175
642	Hypertension complicating pregnancy, childbirth, and the puerperium	3.68	(2.31-5.86)	4.35E-08	90	102,574
420.1	Myocarditis	7.72	(3.69-16.1)	5.43E-08	36	177,003
288.2	Elevated white blood cell count	1.96	(1.54-2.50)	6.43E-08	503	158,244
638	Other high-risk pregnancy	1.85	(1.48-2.31)	8.07E-08	481	178,757
599.3	Dysuria	1.35	(1.21-1.50)	8.61E-08	2383	149,134

727.1	Synovitis and tenosynovitis	1.79	(1.44-2.21)	8.76E-08	688	148,134
276.6	Fluid overload	3.30	(2.13-5.12)	9.52E-08	191	149,733
595	Hydronephrosis	2.45	(1.75-3.44)	1.97E-07	290	176,011
250.2	Type 2 diabetes	1.55	(1.31-1.83)	2.25E-07	1418	148,033
427.7	Tachycardia NOS	1.62	(1.35-1.95)	2.71E-07	857	137,086
38.3	Bacteremia	4.25	(2.44-7.41)	3.32E-07	73	166,009
704.1	Alopecia	1.74	(1.41-2.15)	3.33E-07	563	174,033
649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	3.23	(2.04-5.12)	5.51E-07	96	95,518
260.2	severe protein-calorie malnutrition	3.03	(1.95-4.68)	6.77E-07	165	157,175
790.6	Other abnormal blood chemistry	1.36	(1.21-1.54)	7.51E-07	2209	168,380
535.2	Atrophic gastritis	1.76	(1.40-2.20)	8.30E-07	684	170,481
38	Septicemia	4.68	(2.53-8.68)	9.48E-07	63	166,009
642.1	Preeclampsia and eclampsia	5.74	(2.84-11.6)	1.1E-06	46	102,574
1009	Injury, NOS	1.32	(1.18-1.47)	1.16E-06	2374	148,125
411.2	Myocardial infarction	1.99	(1.51-2.63)	1.24E-06	542	159,985
550.2	Diaphragmatic hernia	1.69	(1.37-2.10)	1.27E-06	787	170,725
306	Other mental disorder	2.43	(1.70-3.49)	1.36E-06	193	124,310
745	Pain in joint	1.24	(1.14-1.35)	1.46E-06	4298	133,022
455	Hemorrhoids	1.41	(1.23-1.62)	1.55E-06	1724	162,711
593	Hematuria	1.57	(1.31-1.89)	1.7E-06	861	152,649
617	Disorders secondary to childbirth, surgery, trauma	1.50	(1.27-1.77)	1.99E-06	934	181,721
994.2	Sepsis	4.06	(2.28-7.25)	2.03E-06	73	178,584
411.4	Coronary atherosclerosis	1.53	(1.29-1.83)	2.1E-06	1374	159,985
647.1	Infections of genitourinary tract during pregnancy	4.11	(2.29-7.36)	2.11E-06	58	102,065
599.2	Retention of urine	1.85	(1.43-2.38)	2.23E-06	527	149,134
272.1	Hyperlipidemia	1.38	(1.21-1.58)	2.96E-06	2322	140,288
567	Peritonitis and retroperitoneal infections	3.53	(2.08-5.99)	2.98E-06	95	166,825
619.4	Noninflammatory disorders of vagina	1.50	(1.26-1.78)	3.04E-06	917	93,150
781	Symptoms involving nervous and musculoskeletal systems	1.92	(1.46-2.53)	3.76E-06	430	180,070
480.2	Viral pneumonia	6.74	(3.00-15.2)	3.78E-06	31	153,134
350.1	Abnormal involuntary movements	1.6	(1.31-1.95)	3.83E-06	807	170,487
502	Postinflammatory pulmonary fibrosis	2.48	(1.68-3.65)	4.54E-06	201	157,792
244.4	Hypothyroidism NOS	1.49	(1.26-1.77)	4.58E-06	1163	160,570
292.4	Altered mental status	2.13	(1.54-2.94)	4.60E-06	315	162,234
656	Other perinatal conditions of fetus or newborn	2.4	(1.65-3.51)	5.26E-06	149	182,164
509.8	Dependence on respirator [Ventilator] or supplemental oxygen	3.24	(1.94-5.41)	7.34E-06	123	157,792
41.9	Infection with drug-resistant microorganisms	6.52	(2.87-14.8)	7.48E-06	30	166,009
401.22	Hypertensive chronic kidney disease	1.98	(1.45-2.69)	1.34E-05	582	122,907
773	Pain in limb	1.23	(1.12-1.36)	1.45E-05	3532	147,087
261.4	Vitamin D deficiency	1.35	(1.18-1.54)	1.77E-05	1782	157,175
395.2	Nonrheumatic aortic valve disorders	1.88	(1.41-2.50)	1.8E-05	412	166,630
1010.6	Reproductive and maternal health services	1.43	(1.21-1.68)	2.04E-05	865	172,787
304	Adjustment reaction	1.47	(1.23-1.76)	2.27E-05	899	124,310
357	Inflammatory and toxic neuropathy	1.51	(1.25-1.83)	2.32E-05	997	174,863
532	Dysphagia	1.46	(1.23-1.75)	2.38E-05	1155	137,590
427.22	Atrial flutter	2.23	(1.54-3.23)	2.40E-05	320	137,086
300.1	Anxiety disorder	1.22	(1.11-1.34)	2.41E-05	3961	124,310
250.42	Other abnormal glucose	1.40	(1.20-1.64)	2.83E-05	1548	148,033

386.9	Dizziness and giddiness (Light-headedness and vertigo)	1.30	(1.15-1.48)	2.84E-05	2066	161,499
250.22	Type 2 diabetes with renal manifestations	2.34	(1.57-3.48)	3.08E-05	496	148,033
587	Kidney replaced by transpant	5.42	(2.44-12.1)	3.38E-05	56	157,475
504	Other alveolar and parietoalveolar pneumonopathy	2.72	(1.69-4.38)	3.56E-05	120	157,792

Table S11. Sensitivity analysis: adults with ≥6 months of EHR records prior to testing.

Phecode	Description	Odds Ratio	95% CI	p value	No. cases	No. controls
512.9	Other dyspnea	2.80	(2.27-3.46)	1.16E-21	680	53,761
512.7	Shortness of breath	2.45	(2.02-2.98)	1.28E-19	825	53,761
569.2	Gastrointestinal complications	7.06	(4.47-11.2)	5.48E-17	87	110,991
278.11	Morbid obesity	2.37	(1.91-2.93)	4.59E-15	506	101,668
136	Other infectious and parasitic diseases	10.0	(5.48-18.4)	8.24E-14	50	123,943
418.1	Precordial pain	3.23	(2.34-4.47)	1.39E-12	254	86,891
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	3.56	(2.42-5.22)	9.13E-11	130	67,497
509.1	Respiratory failure	5.84	(3.36-10.1)	3.71E-10	82	105,765
418	Nonspecific chest pain	1.98	(1.60-2.45)	4.79E-10	607	86,891
585.1	Acute renal failure	3.16	(2.20-4.55)	5.70E-10	258	104,419
427.9	Palpitations	1.98	(1.59-2.47)	1.65E-09	533	87,904
646	Other complications of pregnancy NEC	5.17	(2.90-9.22)	2.50E-08	52	70,931
292	Neurological disorders	2.60	(1.83-3.68)	7.80E-08	209	108,436
782.3	Edema	2.17	(1.63-2.89)	1.01E-07	355	111,634
350.1	Abnormal involuntary movements	2.54	(1.80-3.60)	1.36E-07	213	114,492
781	Symptoms involving nervous and musculoskeletal systems	3.31	(2.10-5.22)	2.43E-07	123	122,421
1010	Other tests	2.88	(1.91-4.34)	4.48E-07	130	112,904
427.21	Atrial fibrillation	2.30	(1.66-3.19)	5.25E-07	335	87,904
644	Anemia during pregnancy	7.11	(3.22-15.7)	1.22E-06	28	73,007
359.2	Myopathy	11.0	(4.16-29.0)	1.34E-06	21	117,651
599.2	Retention of urine	2.97	(1.88-4.69)	2.90E-06	147	94,396
638	Other high-risk pregnancy	1.99	(1.48-2.69)	6.16E-06	254	121,636
502	Postinflammatory pulmonary fibrosis	6.45	(2.79-14.9)	1.31E-05	33	105,765
649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	4.10	(2.14-7.86)	2.06E-05	43	67,497
420.1	Myocarditis	9.91	(3.42-28.7)	2.37E-05	16	120,148
386.9	Dizziness and giddiness	1.66	(1.31-2.10)	2.85E-05	525	105,771
199	Neoplasm of uncertain behavior	2.91	(1.76-4.80)	3.03E-05	107	116,518
1010.6	Reproductive and maternal health services	1.62	(1.29-2.05)	4.50E-05	416	117,591
1013	Asphyxia and hypoxemia	4.90	(2.26-10.6)	5.42E-05	37	119,659

Table S12. Sensitivity analysis: propensity-matched cohort.

Phecode	Description	Odds Ratio	95% CI	p value	No. cases	No. controls
512.9	Other dyspnea	2.88	(2.37-3.50)	3.22E-26	487	53,243
512.7	Shortness of breath	2.35	(1.96-2.82)	1.54E-20	602	53,243
278.11	Morbid obesity	2.16	(1.75-2.65)	2.69E-13	402	103,905
427.9	Palpitations	2.15	(1.74-2.66)	1.27E-12	386	94,868
509.1	Respiratory failure	6.44	(3.73-11.1)	2.48E-11	59	106,519
569.2	Gastrointestinal complications	4.54	(2.88-7.16)	8.01E-11	78	110,663
136	Other infectious and parasitic diseases	10.5	(5.04-21.7)	2.74E-10	39	117,892
418	Nonspecific chest pain	1.89	(1.55-2.31)	4.29E-10	446	92,556
418.1	Precordial pain	2.77	(1.98-3.88)	3.34E-09	149	92,556
649	Conditions of the mother complicating pregnancy, childbirth, or the puerperium	2.95	(2.06-4.21)	3.36E-09	126	62,991
585.1	Acute renal failure	2.96	(2.05-4.26)	6.35E-09	149	106,463
427.21	Atrial fibrillation	2.35	(1.73-3.19)	4.17E-08	209	94,868
359.2	Myopathy	15.7	(5.83-42.0)	4.77E-08	26	114,733
782.3	Edema	2.24	(1.67-3.01)	7.19E-08	200	110,984
646	Other complications of pregnancy NEC	4.67	(2.66-8.22)	8.36E-08	52	65,260
1010	Other tests	2.89	(1.93-4.31)	2.24E-07	100	110,868
1013	Asphyxia and hypoxemia	10.9	(4.40-26.9)	2.4E-07	23	115,135
292	Neurological disorders	2.42	(1.71-3.44)	7.63E-07	139	108,388
1010.6	Reproductive and maternal health services	1.59	(1.31-1.93)	3.29E-06	478	113,604
452.2	Deep vein thrombosis [DVT]	3.05	(1.90-4.90)	3.82E-06	73	107,835
350.1	Abnormal involuntary movements	2.23	(1.59-3.13)	3.99E-06	149	110,095
638	Other high-risk pregnancy	1.94	(1.46-2.58)	4.38E-06	215	116,623
278.1	Obesity	1.7	(1.35-2.14)	6.37E-06	348	103,905
599.2	Retention of urine	2.82	(1.78-4.47)	9.77E-06	82	98,114
514	Abnormal findings examination of lungs	2.19	(1.54-3.12)	1.43E-05	155	108,938
502	Postinflammatory pulmonary fibrosis	9.23	(3.34-25.5)	1.8E-05	17	106,519
285	Other anemias	1.85	(1.39-2.46)	2.19E-05	233	101,673
649.1	Diabetes or abnormal glucose tolerance complicating pregnancy	3.61	(1.98-6.56)	2.64E-05	44	62,991
327.32	Obstructive sleep apnea	1.63	(1.30-2.06)	2.64E-05	368	101,184
781	Symptoms involving nervous and musculoskeletal systems	2.58	(1.63-4.10)	5.69E-05	77	117,346

Table S13. Covariate balance in propensity-matched cohort.

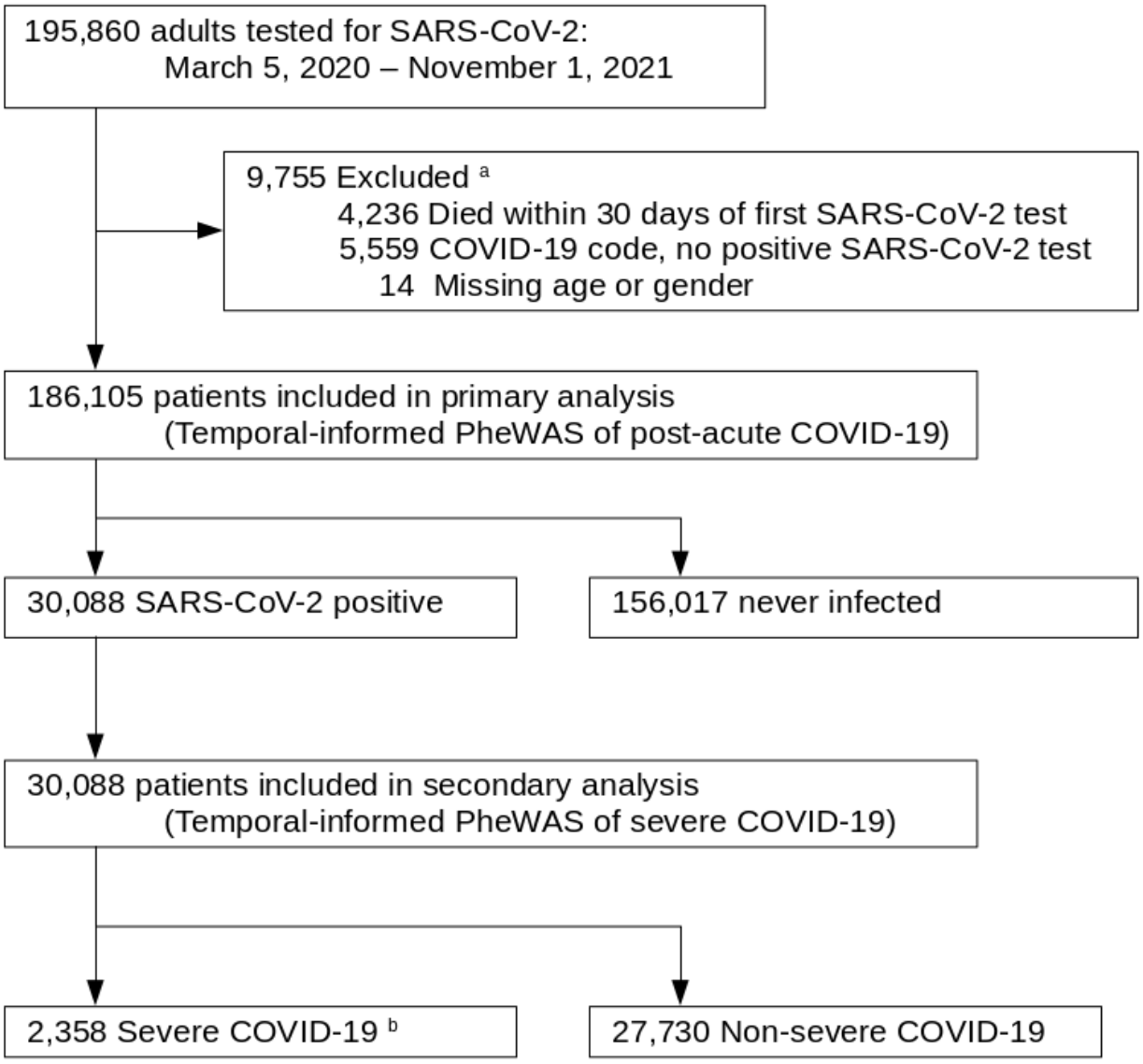
Characteristic	Never Infected	SARS-CoV-2 Positive	Overall	Standardized mean difference[a]
Number in cohort	90,264	30,088	120,352	
Age (median [IQR])	43 [31, 59]	43 [30, 57]	43 [30, 59]	0.004
Sex (%)				0.013
Female	51,747 (57.3)	16,718 (55.6)	68,465 (56.9)	
Male	38,517 (42.7)	13,370 (44.4)	51,887 (43.1)	
Race (%)				
White	60,534 (67.1)	19,176 (63.7)	79,710 (66.2)	0.012
Black	9,654 (10.7)	3,274 (10.9)	12,928 (10.7)	0.027
Other	4,776 (5.3)	1,714 (5.7)	6,490 (5.4)	0.010
Unknown	15,300 (17.0)	5,924 (19.7)	21,224 (17.6)	0.013
Ethnicity (%)				
Non-Hispanic	68,827 (76.3)	21,936 (72.9)	90763 (75.4)	0.010
Hispanic/Latino	3,027 (3.4)	1,217 (4.0)	4244 (3.5)	0.006
Unknown	18,410 (20.4)	69,35 (23.0)	25345 (21.1)	0.013
Received care at VUMC prior to SARS-CoV-2 test (%)	63,028 (69.8)	20,860 (69.3)	83,888 (69.7)	
SARS-CoV-2 testing indication (%)				0.095
Asymptomatic screening	28,982 (32.1)	6,095 (20.3)	35077 (29.1)	
Symptomatic testing	61,282 (67.9)	23,993 (79.7)	85275 (70.9)	
EHR observation time				
Before SARS-CoV-2 test, median [IQR], years	5.6 [0.8, 14.4]	5.8 [0.8, 14.4]	5.7 [0.8, 14.4]	0.001
After recovery, median [IQR], days	393 [242, 509]	361 [285, 427]	381 [252, 500]	0.046
Hospitalization associated with SARS-CoV-2 test (%)	12,434 (13.8)	3,393 (11.3)	15,827 (13.2)	
Time from SARS-CoV-2 test to first follow up (median [IQR])	78 [46, 165]	86 [48, 181]	80 [46, 169]	0.034
Died (%)	531 (0.6)	158 (0.5)	689 (0.6)	0.001
Comorbidites prior to SARS-CoV-2 test (%)				
Myocardial infarction	1,194 (1.3)	419 (1.4)	1,613 (1.3)	0.006
Congestive heart failure	1,811 (2.0)	630 (2.1)	2,441 (2.0)	0.006
Peripheral vascular disease	1,052 (1.2)	325 (1.1)	1,377 (1.1)	0.008
Cerebrovascular disease	1,957 (2.2)	610 (2.0)	2,567 (2.1)	0.01
Dementia	327 (0.4)	137 (0.5)	464 (0.4)	0.015
Chronic pulmonary disease	5,235 (5.8)	1,646 (5.5)	6,881 (5.7)	0.014
Rheumatologic disease	1,555 (1.7)	428 (1.4)	1,983 (1.6)	0.024
Peptic ulcer disease	405 (0.4)	147 (0.5)	552 (0.5)	0.006
Diabetes	4,993 (5.5)	1,964 (6.5)	6,957 (5.8)	0.042
Mild liver disease	2,195 (2.4)	726 (2.4)	2,921 (2.4)	0.001
Severe liver disease	528 (0.6)	170 (0.6)	698 (0.6)	0.003
Hemiplegia or paraplegia	458 (0.5)	153 (0.5)	611 (0.5)	0.001
Renal disease	1,627 (1.8)	750 (2.5)	2,377 (2.0)	0.048
Any malignancy	5,268 (5.8)	1,463 (4.9)	6,731 (5.6)	0.043
Metastatic solid tumor	824 (0.9)	246 (0.8)	1,070 (0.9)	0.01
AIDS or HIV infection	606 (0.7)	173 (0.6)	779 (0.6)	0.012

Variables used to develop the propensity-scoring model for probability of testing positive for SARS-CoV-2 included age, sex, race, ethnicity, symptomatic testing indication, inpatient hospitalization around time of SARS-CoV-2 test, observation time after recovery, and length of EHR prior to SARS-CoV-2 testing.

[a] Standardized mean difference values of less than 0.1 indicate acceptable matching between groups.

Supplementary Figures

Figure S1. Study CONSORT diagram

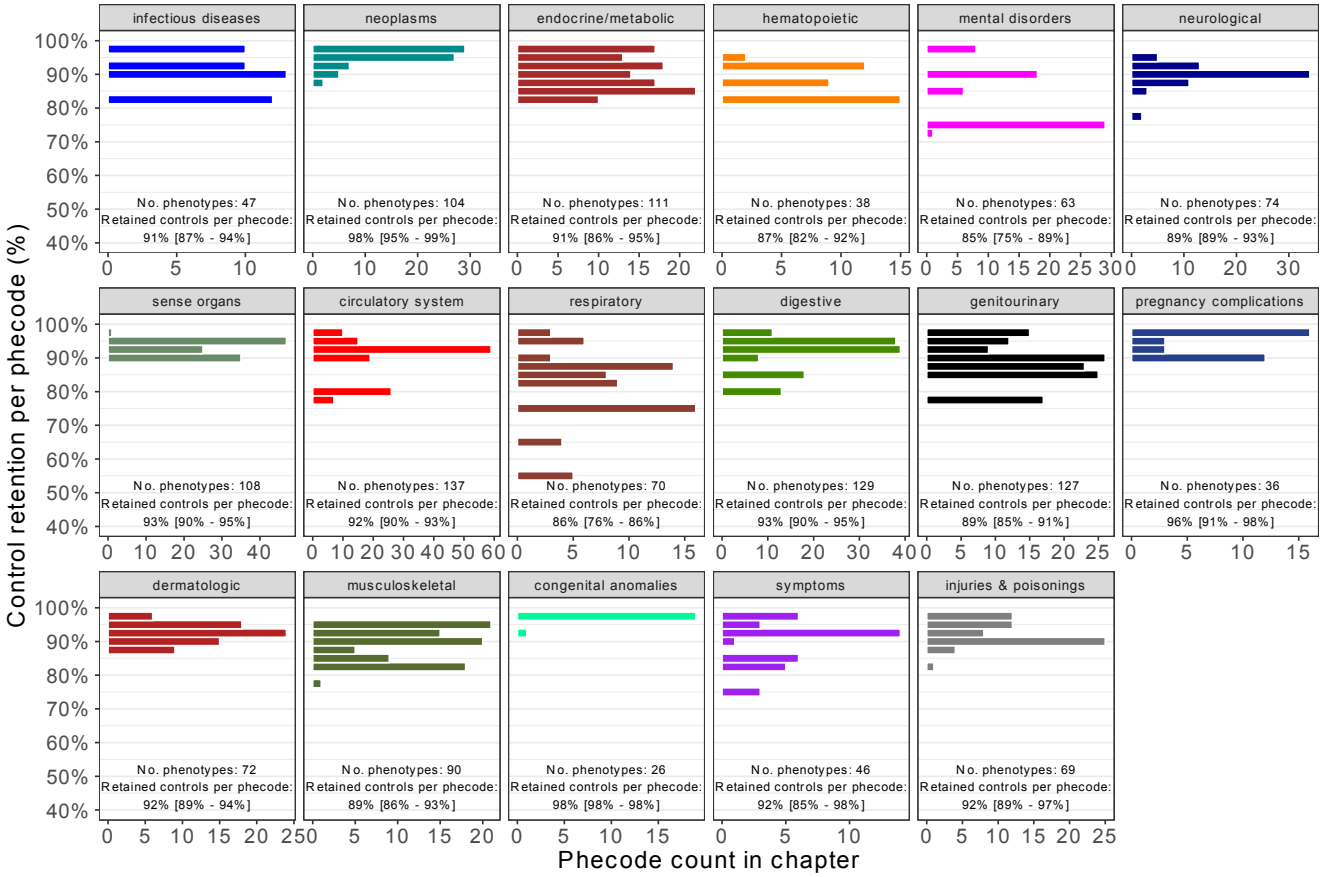


Flow diagram of adult patients in Vanderbilt COVID-19 EHR registry database, patients excluded, and numbers included in analyses. SARS-CoV-2: severe acute respiratory syndrome coronavirus 2. COVID-19: Coronavirus disease 2019.

[a] Some patients had more than one reason for exclusion.

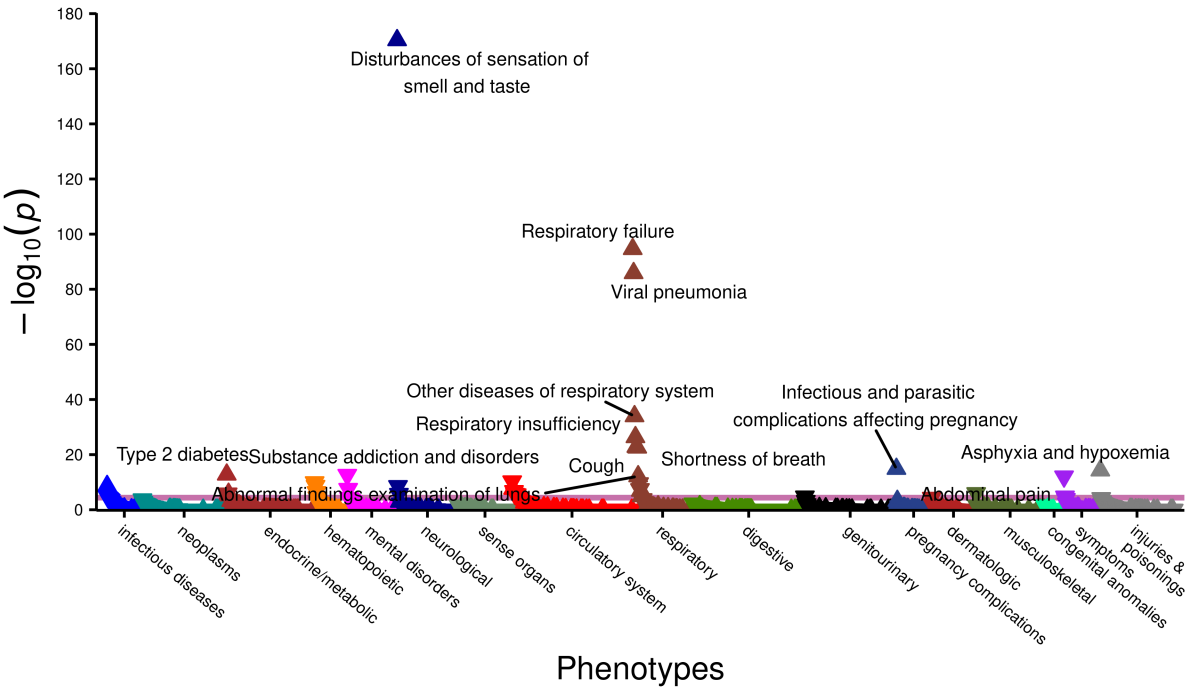
[b] Severe COVID-19: admitted to hospital and required supplemental oxygen.

Figure S2. Control retention by temporal-informed phenotyping



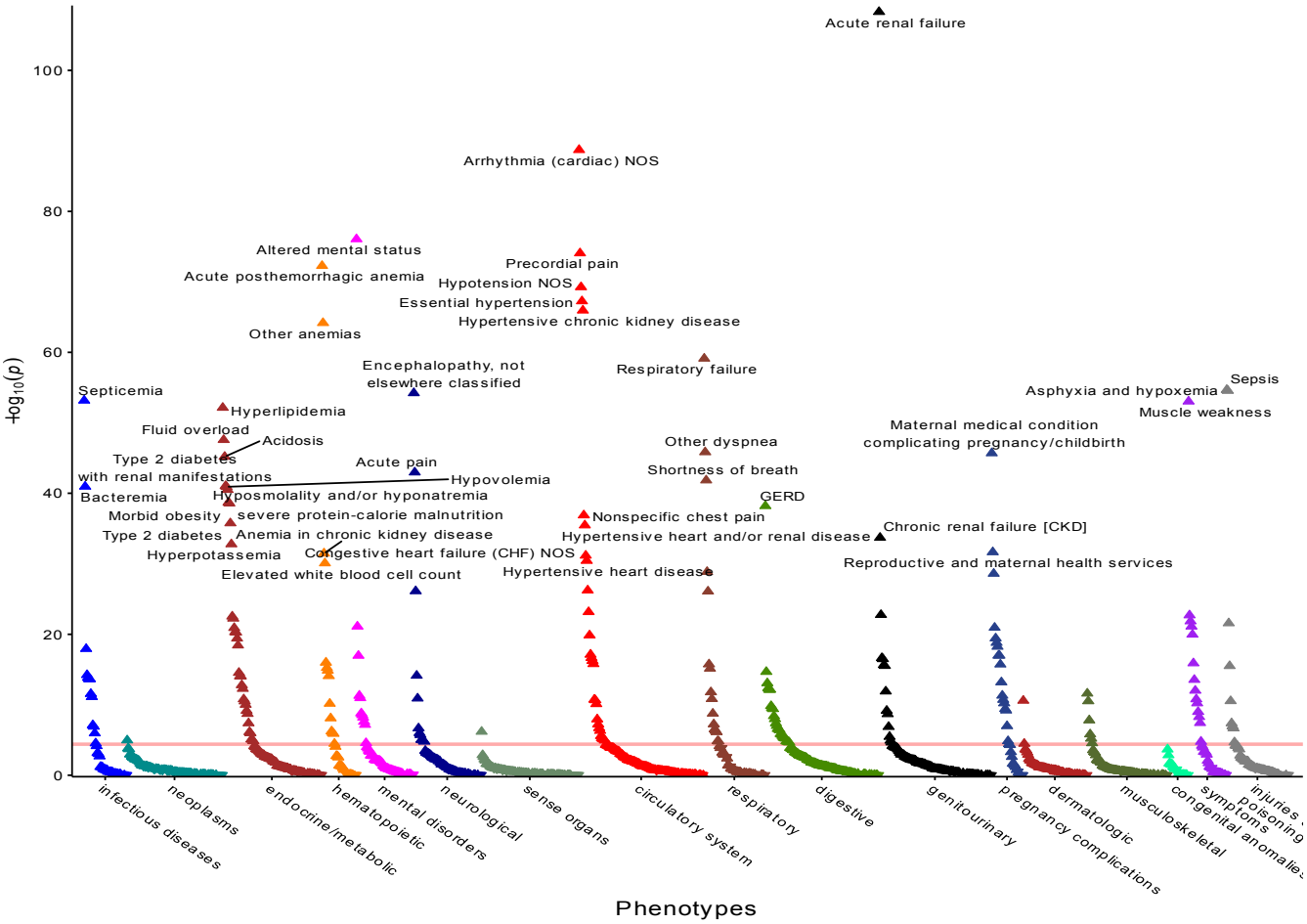
Histograms of phenotype control retention per PheWAS code (phecode) using temporal-informed phenotyping. Individual histograms indicate each chapter within the phecode hierarchy.[32] Number of phecodes per chapter are shown on *x* axis, control retention per phecode is shown on *y* axis,. Labels indicate number of phenotypes with ≥10 cases and median [interquartile range] of the per-phecode case retention in each chapter.

Figure S3. Phenome scan of phenotypes occurring around date of SARS-CoV-2 testing



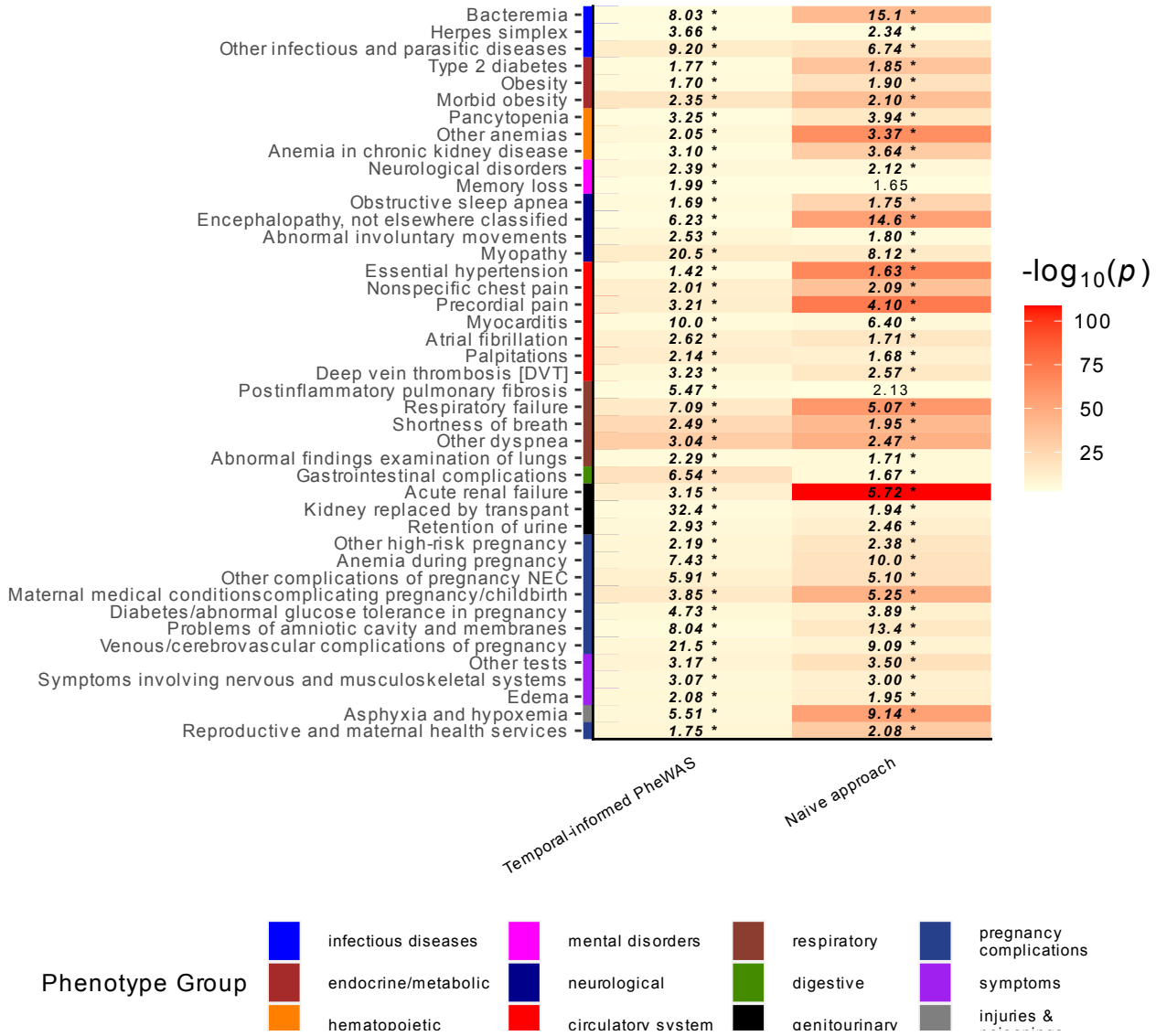
PheWAS plot of medical phenotypes associated with a positive SARS-CoV-2 test for phenotypes entered into the EHR within ±15 days of the index test date (n = 186,105, phenotypes available for testing = 1,182). The *x* axis represents phecodes grouped by chapter within the phecode hierarchy.[10] The *y* axis represents the negative log-transformed *p* values obtained using logistic regression after adjusting for age, sex, race, ethnicity, length of EHR observation after recovery, indication for testing, and medical comorbidities prior to testing. Upward triangles represent phenotypes with odds ratio >1.0 for COVID-19 survivors and downward triangles represent phenotypes with odds ratio <1.0.

Figure S4. Phenome scan using naive post-acute phenotyping.



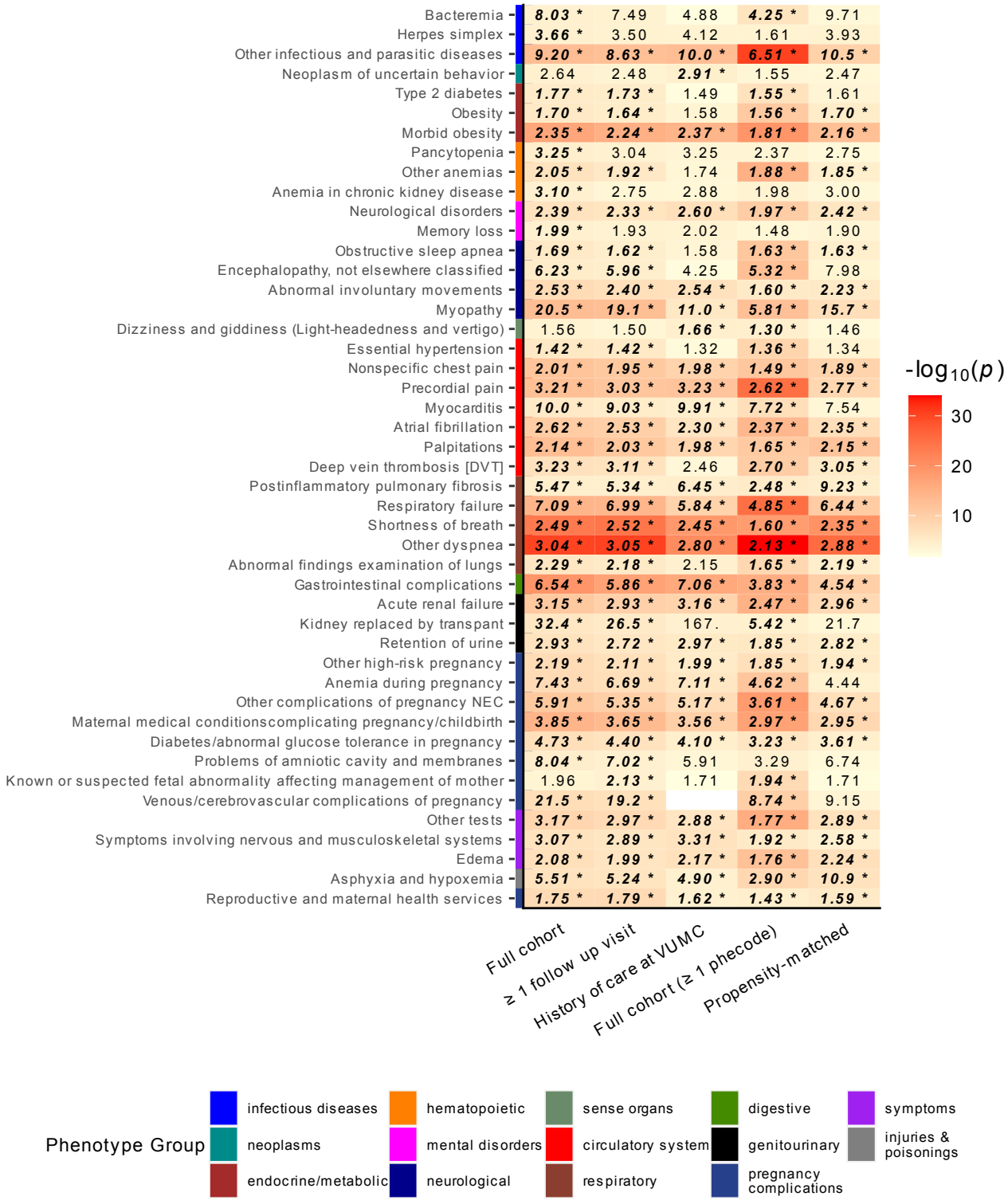
PheWAS plot of new post-acute phenotypes identified by naive phenotyping (all post-acute codes) for COVID-19 survivors versus never-infected patients as the referent group (n = 186,105, phenotypes available for testing = 1,347). The *x* axis represents phecodes grouped by chapter within the phecode hierarchy. The *y* axis represents the negative log-transformed *p* values obtained using logistic regression after adjusting for age, sex, race, ethnicity, length of EHR observation after recovery, indication for testing, and medical comorbidities prior to testing. Upward triangles represent phenotypes with odds ratio >1.0 for COVID-19 survivors and downward triangles represent phenotypes with odds ratio <1.0. Horizontal red line indicates the phenome-wide significance *p* value significance using a Bonferroni correction.

Figure S5. Comparison of PheWAS results by temporal-informed or naive phenotyping.



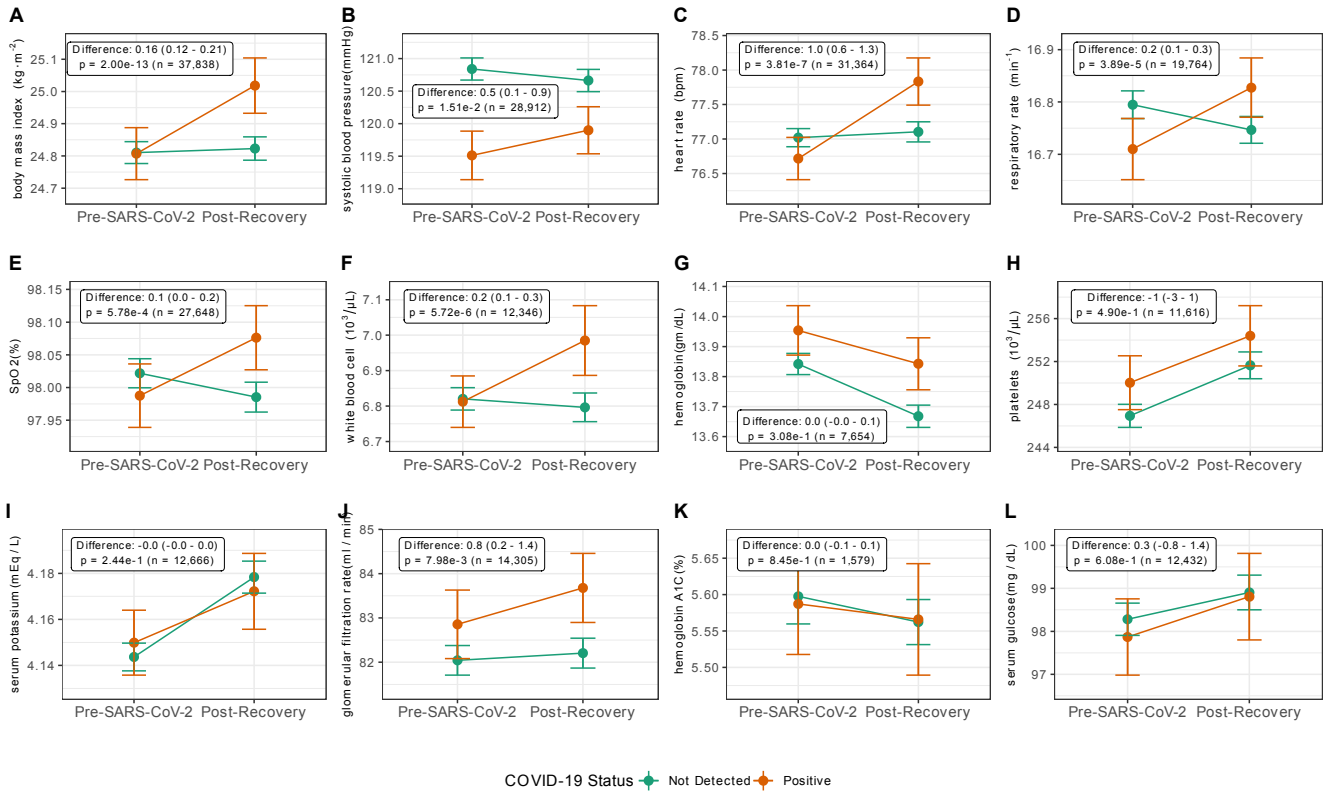
Comparison of PheWAS results using temporal-informed phenotyping (left column) and naive post-acute phenotyping (right column). The *y* axis represents phenotypes (as PheWAS codes / "phecodes") that are group by category within the phecode hierarchy.[10] Cell color intensity illustrates adjusted *p* values by logistic regression. Text in cells show point estimates for effect odds ratios. Text in bold/italic and with a '*' indicate PheWAS associations that were statistically significant using a Bonferroni correction. Results for phecodes that were significant in the primary analysis (left column) are displayed for brevity. Results in bold/italic text and with a '*' indicate PheWAS associations that were statistically significant using a Bonferroni-corrected *p* value.

Figure S6. Temporal-informed PheWAS: sensitivity analyses.



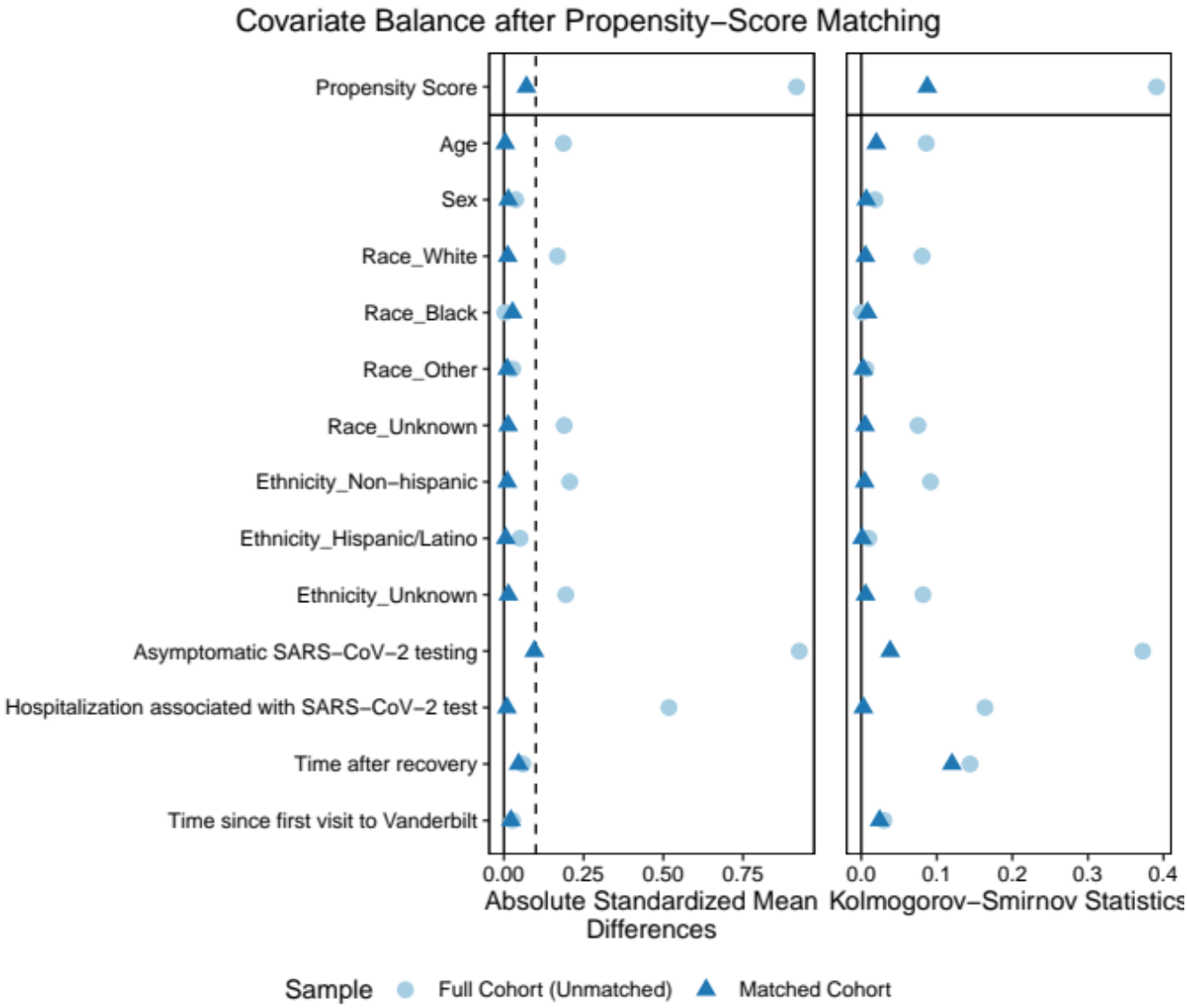
PheWAS results for new post-acute phenotypes identified during outpatient follow-up among all adults tested for SARS-CoV-2 (left column, n=186,105), and using different sensitivity analyses including only patients who had a subsequent follow-up visit at our center (2nd column, n = 113,198, phenotypes available for testing = 909); (2) patients with a history of care at our institution (3rd column, n = 127,699, phenotypes available for testing = 845); using a single phecode as the case definition rather than the standard two phecodes on separate dates (4th column, n = 186,105, phenotypes available for testing = 1,309); and a propensity-matched cohort (5th column, n = 120,352, phenotypes available for testing = 819). The *y* axis represents phenotypes (as PheWAS codes / "phecodes") that are group by category within the phecode hierarchy.[10] Cell color intensity illustrates adjusted *p* values by logistic regression. Text in cells show point estimates for effect odds ratios. Text in bold/italic and with a '*' indicate PheWAS associations that were statistically significant using a Bonferroni correction. Results for phecodes that were significant in the primary analysis (left column) are displayed for brevity. Results in bold/italic text and with a '*' indicate PheWAS associations that were statistically significant using a Bonferroni-corrected *p* value. Empty cells indicate analyses with insufficient phenotype cases (less than 10) to perform the analysis for that phenotype.

Figure S7. Changes in vital signs and laboratory test values: all assessed labs.



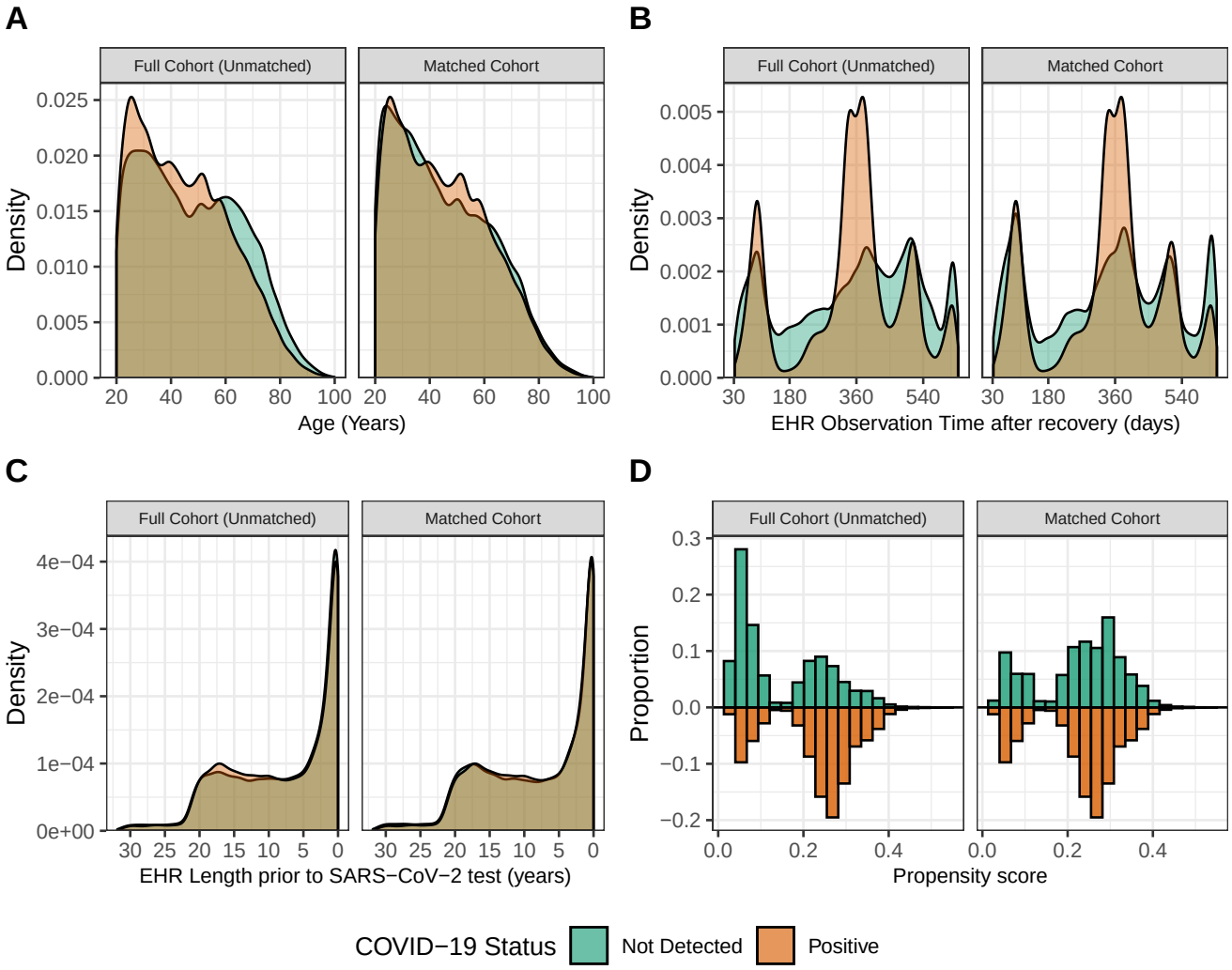
Changes in vital signs and laboratory tests between pre-testing and post-recovery periods among COVID-19 survivors (orange) versus never-infected controls (green). For each patient we used the median pre-testing values obtained during outpatient visits occurring within 180 days before the index SARS-CoV-2 test, and the median post-recovery values obtained during outpatient visits occurring within 365 days after recovery from illness. Dots represent mean values in each exposure group, bars represent standard errors of the mean. Labels represent the adjusted mean difference between COVID-19 survivors and never-infected controls, number of patients with data for each analysis, and p-values obtained by multiple linear regression. (**A**) Body mass index, (**B**) systolic blood pressure, (**C**) heart rate, (**D**) respiratory rate, (**E**) oxygen saturation by pulse oximetry, (**F**) white blood cell count, (**G**) hemoglobin, (**H**) platelet count, (**I**) serum potassium, (**J**) estimated glomerular filtration rate, (**K**) hemoglobin A1c, and (**L**) serum glucose.

Figure S8. Covariate balance with propensity-score matching.



Love plot summarizing of covariate balance before (light blue dots, n=186,105) and after propensity-score matching (blue triangles, n=120,352). The x-axis indicates balance statistic value. The y-axis represents the overall propensity score (above black horizontal line) and each conditioning variable in the propensity scoring model (below black horizontal line). Left graph illustrates absolute standardized mean differences. Right graph illustrates the Kolmogorov-Smirnov statistic (maximum difference in empirical cumulative density function). For the absolute standardized mean difference, a threshold of 0.1 (dashed vertical line) is indicative of good matching between the COVID-19 and never-infected groups. For the Kolmogorov-Smirnov statistic, no specific threshold is recommended but improvements (values closer to zero) are noted for all variables after matching.

Figure S9. Covariate balance with propensity-score matching for continuous variables.



Density plots of **A.** age, **B.** EHR observation time after recovery (in days), **C.** length of EHR record prior to testing (in years), and **D.** histogram of propensity scores both before (full unmatched cohort, n=186,105) and after (n=120,352) propensity matching. For all variables, we observed an improvement in matching between COVID-19 survivors (orange) and never-infected controls (green). The differences in distributions in observation time **B.** reflect the higher positive test rates during local waves of the pandemic, whereas the distribution of negative tests was more uniform over the course of the pandemic. Propensity matching resulted in acceptable balance for all variables.

Supplementary References

1 Danciu I, Cowan JD, Basford M, *et al.* Secondary Use of Clinical Data: the Vanderbilt Approach. *J Biomed Inform* 2014;**52**:28–35. doi:10.1016/j.jbi.2014.02.003

2 FitzHenry F, Resnic FS, Robbins SL, *et al.* Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership. *Appl Clin Inform* 2015;**6**:536–47. doi:10.4338/ACI-2014-12-CR-0121

3 DeLozier S, Bland S, McPheeters M, *et al.* Phenotyping coronavirus disease 2019 during a global health pandemic: Lessons learned from the characterization of an early cohort. *J Biomed Inform* 2021;**117**:103777. doi:10.1016/j.jbi.2021.103777

4 Denny JC, Giuse DA, Jirjis JN. The Vanderbilt Experience with Electronic Health Records. *Semin Colon Rectal Surg* 2005;**16**:59–68. doi:10.1053/j.scrs.2005.08.003

5 Feng Q, Wei W-Q, Chaugai S, *et al.* Association Between Low-Density Lipoprotein Cholesterol Levels and Risk for Sepsis Among Patients Admitted to the Hospital With Infection. *JAMA Netw Open* 2019;**2**:e187223. doi:10.1001/jamanetworkopen.2018.7223

6 Quan H, Sundararajan V, Halfon P, *et al.* Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Medical Care* 2005;**43**:1130–9. doi:10.1097/01.mlr.0000182534.19832.83

7 Denny JC, Ritchie MD, Basford MA, *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics* 2010;**26**:1205–10. doi:10.1093/bioinformatics/btq126

8 Ho D, Imai K, King G, *et al.* MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* 2011;**42**:1–28. doi:10.18637/jss.v042.i08

9 Greifer N. *cobalt: Covariate Balance Tables and Plots*. 2022. https://ngreifer.github.io/cobalt/

10 Wu P, Gifford A, Meng X, *et al.* Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform* 2019;**7**:e14325. doi:10.2196/14325

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	**Title and Abstract.**	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	RECORD 1.1: **Abstract Methods** RECORD 1.2: **Title, Abstract Methods** RECORD 1.3: Not applicable.
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	**Introduction**.		
Objectives	3	State specific objectives, including any prespecified hypotheses	**Introduction**.		
Methods					
Study Design	4	Present key elements of study design early in the paper	**Methods**. **2.1, 2.2, 2.3, 2.4**		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	**Methods 2.1, 2.2 Supplementary Methods**		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	**(a) Methods 2.1, 2.2, 2.3**	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.	RECORD 6.1: **Methods 2.1**

		Case-control study - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	RECORD 6.2: **Supplementary Methods** RECORD 6.3: Not applicable.
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.	Outcomes: **Methods 2.4** Exposures: **Methods 2.5** Confounders / covariates. **Supplementary Methods**	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	RECORD 7.1. **Supplementary Methods** and **Tables S2**
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	**Supplementary Methods**		
Bias	9	Describe any efforts to address potential sources of bias	**Methods 2.5 Supplementary Methods**		
Study size	10	Explain how the study size was arrived at	**Methods 2.1**		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	**Methods 2.5, Supplementary Methods**.		

Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	**(a), (b) Methods 2.5** **(c), (d) Supplementary Methods**		
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	RECORD 12.1: **Supplementary Methods** RECORD 12.2: **Supplementary Methods**
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	RECORD 12.3: Not applicable
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the	**(a) Results 3.1** **(b) Results 3.1** **(c) Results 3.1, Figure S1**	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection	RECORD 13.1 **Results 3.1**, **Figure S1**, **Supplementary Methods**

		study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		of included persons can be described in the text and/or by means of the study flow diagram.	
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	**(a,b,c) Results 3.1, Table 1.**		
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	**Results 3.3, 3.4, 3.5**.		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	**Results 3.3, 3.4, 3.5.** **Tables 3-5** **Figures 3, 5**		

Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	**Results 3.3. Subgroup analyses: Tables S6-S8, Figure 4 Sensitivity analyses: Tables S9-S12, Figure S6**		
Discussion					
Key results	18	Summarise key results with reference to study objectives	**Discussion 4.1**		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	**Discussion 4.4**.	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	RECORD 19.1 **Discussion 4.4**
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	**Discussion 4.1, 4.2, 4.3**		
Generalisability	21	Discuss the generalisability (external validity) of the study results	**Discussion 4.4**		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	**Methods 2.6, Acknowledgments, Declaration of interests**		
Accessibility of protocol, raw data, and programming code				RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	**Data sharing.**

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.