The Coviral Portal: Multi-Cohort Viral Loads and Antigen-Test Virtual Trials for COVID-19

Alexandra Morgan,[1] Elisa Contreras, [1] Michie Yasuda, [1] Sanjucta Dutta, [1] Donald Hamel,[3] Tarini Shankar,[1,4] Diane Balallo,[1,5] Stefan Riedel, [1,2] James E. Kirby, [1,2] Phyllis J. Kanki,[3] and Ramy Arnaout[1,2,6]†*

[1]Beth Israel Deaconess Medical Center, Division of Clinical Pathology, Department of Pathology, Boston, MA 02215, USA; [2]Harvard Medical School, Boston, MA 02215, USA; [3]Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA; [4]Northeastern University, Boston, MA 02215, USA; [5]Boston College, Chestnut Hill, MA 02467, USA; [6]Beth Israel Deaconess Medical Center, Division of Clinical Informatics, Department of Medicine, Boston, MA 02215, USA

†Corresponding author and

*Lead Contact:

Ramy Arnaout

Beth Israel Deaconess Medical Center

330 Brookline Avenue, DA709

Boston, MA 02215

rarnaout@bidmc.harvard.edu

617-538-5681

Abstract

Background. Regulatory approval of new over-the-counter tests for infectious agents such as SARS-CoV-2 has historically required that clinical trials include diverse groups of specific patient populations, making the approval process slow and expensive. Showing that populations do not differ in their viral loads—the key factor determining test performance—could expedite the evaluation of new tests.

Methods. 46,726 RT-qPCR-positive SARS-CoV-2 viral loads were annotated with patient demographics and health status. Real-world performance of two commercially available antigen tests was evaluated over a wide range of viral loads. An open-access web portal was created allowing comparisons of viral-load distributions across patient groups and application of antigen-test performance characteristics to patient distributions to predict antigen-test performance on these groups.

Findings. In several cases distributions were surprisingly similar where a difference was expected (e.g. smokers vs. non-smokers); in other cases there was a difference that was the opposite direction from expectations (e.g. higher in patients who identified as White vs. Black). Sensitivity and specificity of antigen tests for detecting contagiousness were similar across most groups. The portal is at https://arnaoutlab.org/coviral/.

Conclusions. In silico analyses of large-scale, real-world clinical data repositories can serve as a timely evidence-based proxy for dedicated trials of antigen tests for specific populations. Free availability of richly annotated data facilitates large-scale hypothesis generation and testing.

Funding. Funded by the Reagan-Udall Foundation for the FDA (RA and JEK) and via a Novel Therapeutics Delivery Grant from the Massachusetts Life Sciences Center (JEK).

Keywords: SARS-CoV-2, COVID-19, RT-qPCR, viral load, antigen test, precision medicine, comorbidities, big data, real-world data

Introduction

Diagnosis of new infectious pathogens such as SARS-CoV-2 requires development of new diagnostic tests, which must be evaluated and approved by regulatory agencies before they can be used for patient care. Such tests include over-the-counter (OTC) antigen tests, which have been widely used for at-home testing in the context of COVID-19. In order to be approved, a new test must demonstrate a minimum level of clinical performance. Performance is typically measured as the test's sensitivity, defined as the proportion of true-positive samples that have a positive result, and its specificity, defined as the proportion of true negatives that have a negative result. Clinical performance must be demonstrated in a defined patient population or group and clinical context, for example inpatients as opposed to outpatients. However, at the start of an outbreak, epidemic, or pandemic, there may not be enough information to know whether a test can be expected to perform differently in some patient groups vs. others. Therefore at the start of development, a new diagnostic test may be approved based on its performance in the general population, not specific groups.

In contrast, as time goes on, evidence for clinical differences among specific groups may emerge. As this happens, it becomes reasonable to ask whether a test might perform differently in specific groups, with important implications for how and potentially even whether that test should be used in a given clinical scenario. Ideally, this question would be answered by conducting dedicated trials of the new diagnostic test in each group of interest. Unfortunately, trials are expensive and slow. Also, the number of specific groups that may be of interest is large, since specific subgroups can be defined not only based on demographics (such as age, race, and gender), comorbidities (such as diabetes, heart disease, or immunosuppression), and care settings (inpatient vs. outpatient vs. emergency room) but also by the many possible combinations of each of these characteristics, which is an essential component of precision medicine.[1] As a result, in practice it is prohibitively difficult to perform many separate trials on specific groups for even a single diagnostic test, much less for the many tests that are likely to be developed in response to a large-scale outbreak, such as has happened in response to the COVID-19 pandemic. This is a problem for regulators, clinicians, and patients alike.

One solution is to apply a new test's various performance characteristics to real-world data collected in the course of patient care. Such characteristics include results of existing trials as well as analytical (i.e. pre-clinical) operating parameters such as the limit of detection (LOD). The LOD is defined as the lowest concentration of virus that the test can detect in 95% of replicates. It is routinely determined by manufacturers and validated by clinical laboratories before a test is put to use clinically.[2] The relationship between concentration and detection is usually understood to follow an S-shaped curve;[3] fitting it requires at least one additional datapoint besides the LOD. The concentration of the virus may be measured as the viral load, most often defined as the number of copies of viral mRNA per milliliter of testing material (copies/mL).

"Real-world data" means the viral-load result of a reference diagnostic test that has already been approved for the general population. Because this data is from the general population, it will presumably include results on many specific patient groups. One can apply the performance of the new test as described above to the set or "distribution" of viral loads from a group to predict what proportion of patients in the group would have tested positive with the new test. This proportion is the sensitivity of the new test for that group. In this way, one can estimate clinical sensitivity without needing a dedicated trial on that group.

In this study, we apply this approach to COVID-19. We use the 46,726 positive SARS-CoV-2 RT-qPCR results our institution has performed as of this writing and use our electronic health record to annotate each result according to the patient's demographics, comorbidities, and so on. Importantly, we convert each PCR result from a Ct value to a viral load using robust (100% code-coverage) and accurate publicly available software, as previously described.[2,4–6] Although PCR results are typically reported simply as positive or negative, qPCR is intrinsically quantitative (the "q" in "qPCR"); we make use of this quantitative information in its natural unit of measure (viral load). This is in contrast to Ct values, which are less useful because they vary inversely with viral load and correspond to different viral loads on different PCR testing platforms.

We focus especially on sensitivity and specificity for infectivity or contagiousness. Contagiousness is of special interest given the public health focus on using quick, inexpensive tests to curtail community transmission in a pandemic.[7,8] Treating contagiousness as a function of viral load, contagiousness can be estimated using a virus culture assay we previously described in which a positive patient sample is applied to susceptible cells and monitored for virus replication. After an initial adsorption period, the cells are washed free of the initial virus to eliminate carryover. The supernatant is then sampled on a timescale of days and tested by PCR for the presence of new virus.[9,10] The lowest concentration of virus in a patient sample from which new virus can be recovered is the contagiousness threshold. Because cells in culture have no physical or distance barriers, no mucociliary elevator, and no protection via medications or an immune system, we consider this threshold a conservative estimate. We previously demonstrated this threshold is approximately 50,000 copies/mL and has been fairly stable even as the SARS-CoV-2 virus has evolved.[9,10]

Materials and Methods

Institutional review. Institutional Review Board approval was obtained for all described work under Beth Israel Lahey Health (BILH) IRBs 2022P000328 and 2022P000288. The Harvard T. H. Chan School of Public Health IRB20-1979 provided non-human subjects research determination for virus culture work.

Defining specific patient groups. We extracted the following information from our hospital's clinical-research data repository: demographics (age, gender, and self-reported race/ethnicity), socioeconomic status (using the median neighborhood household income for the patient's ZIP code, obtained via the 2020 U.S. census, as a proxy), care setting (inpatient, outpatient, emergency ward, or other institution), presentation/disposition (based on vital signs, which we combined into a measure of initial presentation), outcome (survived, died with COVID-19 as the cause of death, died with COVID-19 as an incidental finding), vaccination status (vaccinated, unvaccinated, or unknown), treatment (CPT-encoded procedures, remdesivir (GS-5734; Gilead Sciences, Foster City, CA) administration, steroid administration), comorbidities (according to

the Charlson Comorbidity Index:[11] body-mass index, diabetes, chronic heart disease, chronic lung disease, chronic renal disease, liver disease, dementia, chronic neurological conditions, connective-tissue disease, human immunodeficiency virus (HIV), and malignancy), and immunosuppression status[12] (CD4+ T-count <100 cells/µL, hematologic malignancy, chemo/immuno-modulating agent alone or in setting of solid malignancy, organ transplant, or rheumatologic/inflammatory condition). The rationale for extracting these data items specifically was twofold: first, this list includes the complete COVID-19 core diagnostic data at federal and state levels; second, it includes data necessary for calculating the well validated 4C Mortality Score for SARS-CoV-2.[13] ICD-10 codes corresponding to the listed comorbidities were determined by a physician (Dr. Arnaout) following prior methodologies[14] but updated for 2022-2023.

At presentation, patients were considered sick if any of the following were true within 1 day of the PCR test sample: systolic blood pressure <90 mmHg, diastolic blood pressure <60 mmHg, heart rate >100 beats per minute, respiratory rate >18 breaths per minute, or temperature >99.1°F. They were otherwise considered well, with the exception that if no values were recorded (NULL in the data repository) for all criteria, presentation was considered unknown and therefore not assigned.

Patients were designated as immunocompromised at the time of PCR testing if one of the following were true: on their most recent T-cell subset analysis report, their absolute CD4+ cell count was <100 cells/µl; they had a diagnosis of either lymphoma or leukemia associated with a healthcare encounter (visit, admission, or phone call) either before the PCR test or within 60 days after the PCR test; they were on any of the following medications on an ongoing basis, prescribed prior to the PCR test and with enough refills to include the time up to 30 days prior to the PCR test: abatacept, adalimumab, anakinra, azathioprine, basiliximab, budesonide, certolizumab, cyclosporine, daclizumab, dexamethasone, everolimus, etanercept, golimumab, infliximab, ixekizumab, leflunomide, lenalidomide, methotrexate, mycophenolate, natalizumab, pomalidomide, prednisone, rituximab, secukinumab, serolimus, tacrolimus, tocilizumab,

tofacitinib, ustekinumab, or vedolizumab. Otherwise, they were designated not immunocompromised.

Supplementary Table 1 provides further details for the above methods.

Viral load. The SARS-CoV-2 RT-qPCR testing in this study was performed on three Abbott Molecular platforms: m2000, Alinity m, and Alinity 4-Plex (Abbott Molecular, Des Plaines, IL, U.S.A.). These detect identical SARS-CoV-2 N and RdRp gene targets. They are extremely sensitive, with LOD of ~100 copies/mL. They output a quantitative fractional cycle number (FCN), a type of Ct value described in detail elsewhere.[15] Together these platforms accounted for 46,726 positive tests.

Ct values were converted to viral loads in units of copies of viral mRNA per mL using the public Python package *ct2vl* as previously reported.[4] Briefly, this software was validated via calibration curves established for all platforms using an extended SeraCare panel (LGC Seracare, Milford, MA) panel based on a SARS-CoV-2 genome incorporated into replication-incompetent, enveloped Sindbis virus and calibrated based on digital PCR at US National Institutes of Standards and Technology (NIST) and LGC/Seracare.[16] Validation material ranged in viral load from 300 to 10^6 viral genome copies/mL. Results were harmonized with the cycle threshold for a spiked internal control also amplified in each SARS-CoV-2 assay to confirm lack of PCR inhibition and accurate viral load output. The standards, modeling SARS-CoV-2 virus, were run through all stages of sample preparation and extraction to allow appropriate comparison with identically processed patient samples. R^2 was ~0.99 for all calibration determinations, indicating assays are robustly quantitative.

Presumed SARS CoV-2 variant. Presumed variant was inferred from the date of sample collection based on the data presented by Covariants (https://covariants.org) showing the frequency of sequencing particular variants in Massachusetts, the United States, and other locations.[17] Specimens from before June 7, 2021 were annotated as being an early variant. Specimens from between July 7, 2021 and December 6, 2021 were annotated delta variant.

Specimens from after January 3, 2022 were annotated as omicron variant. Results from the month between windows, when more than one major variant was common, were not annotated with a presumed variant and are omitted from by-variant comparisons.

Evaluation of antigen tests vs. PCR. Patients seeking COVID testing at a drive-through testing site near Boston affiliated with our medical center[5,6] between May 23 and November 4 of 2022 were offered the opportunity to participate in a separate arm, providing a comparative, parallel prospective study. This represented community testing for both symptomatic and asymptomatic individuals with diverse demographics (age, race, sex, socio-economic status). Each patient who consented had both a standard-of-care PCR test and two OTC antigen tests performed (Abbott BinaxNow COVID-19 Ag Card and CareStart COVID-19 Antigen Home Test). The PCR test was performed on material collected with a nasopharyngeal swab. SARS-CoV-2 RT-qPCR testing was performed using the Abbott m2000 RealTime or Alinity m SARS-CoV-2 assays according to the manufacturer's instructions, yielding, for each positive sample, a Ct value which was converted to viral load as previously described. Specimens for the antigen tests were collected with separate nasal swabs for each test, according to the manufacturer's instructions. These were collected and the tests performed by study personnel after informed consent was obtained on-site within the time-frame constraints detailed in each test's instructions for use, as per IRB. In order to extrapolate antigen-test performance from this subset to all patients, positivity vs. viral load was modeled by logistic regression (the LogisticRegression function in Python's scikit-learn library).[18] LogisticRegression converges on optimal parameters in a model predicting the probability of a positive test based on viral load. Parameters were predicted separately for each test. The equation for probability was a standard sigmoid constrained to the range 0-1 (i.e., the lowest probability is zero and the highest probability is 1): p(test success) = $\frac{1}{1+e^{-k(v-v_0)}}$ where v, the independent variable, is $\log_{10}$ of the viral load. This constraint leaves two free parameters: v_0 is the midpoint, i.e. the model's estimate of where the success rate passes 50%, while k controls the steepness, i.e. the change in viral load to change in probability of being positive.

Contagiousness. Samples were then stored at 4°C until contagiousness testing. This was done within a four-day time period. We previously validated that freeze-thaw does not impact viral viability and will bank remaining samples for future investigations. Quantitative viral culture was performed on a random sample of PCR-positive samples. Vero E6 cells (ATCC CRL-1586) were seeded on a 6-well flat bottom plate at 0.3×10^6 cells per well in Eagle's minimum essential media (EMEM) containing 1% antibiotic-antimycotic, 1% HEPES and 5% fetal calf serum (FCS, Gibco) grown to confluence at approximately 1×10^6 cells per well, inoculated with 250µL of patient sample, and incubated at 37°C for 24 hours for viral adsorption, as previously described.[10,19–21] Carryover of non-viable viral RNA present in samples was limited by washing cell cultures after the 24-hour viral adsorption and adding fresh EMEM composite media with reduced FCS to 2% for viral growth, meaning detectable virus represents viable replicating virus. On days 3 and 6, cell culture supernatant was removed and added to 800µL of VXL buffer (QIAGEN, German, MD) (1:1 ratio) for subsequent nucleic acid extraction and detection of virus by PCR. Viral load in culture supernatants on days 3 and 6 served as a quantitative surrogate for viable (i.e. replication-competent) virus in the patient sample and provided a measure of the magnitude of sample infectivity. SARS-CoV-2 RT-qPCR testing of Vero cell culture supernatants was performed using the Abbott m2000 Real-Time or Alinity m SARS-CoV-2 assays according to the manufacturer's instructions. The contagiousness threshold was determined by the threshold patient-sample viral load value resulting in detectable culture viral load.

Whole-genome viral NGS. Next-generation-sequencing (NGS)-based sequencing of select PCR-positive samples from the viral antigen evaluation study was performed as follows. Full-length SARS-CoV-2 viral genome sequencing was performed on the Oxford Nanopore MinION system (≥R9.4 flowcell; Oxford Nanopore Technologies-ONT, Oxford, UK) using the guppy basecaller and the downstream ARTIC network bioinformatics pipeline for genome assembly.[22,23] The workflow was run on a 2021 Intel Core i9-11900 Rocket Lake 3.5GHz 8-core LGA 1200 boxed processor with NVIDIA A5000 GPU. Standard coverage and quality metrics and plots were produced, single-nucleotide variants were recorded, and variants assigned using NextClade.[24]

Web portal and privacy protection. The portal was written using Svelte and d3 for the interactive frontend and Python run against a Postgres database for the backend. To reduce re-identification risk, ages were jittered by adjusting the patient's date of birth by a random number of days (drawn from a Gaussian distribution with a standard deviation of two years) before calculating patient's age at the time of each test. Groups smaller than 4-8 patients are suppressed and therefore not viewable. Revealing exact sizes of such small groups defined by multiple patient characteristics would pose a re-identification risk. To prevent inferring the sizes of these groups by subtraction of viewable group sizes, viewable group sizes are jittered by dropping approximately 0.5-1% of the data on any split by patient feature. To maximize consistency of the results of jittering as data are updated, jittering was performed using random number seeds based on pseudo-identifiers (which are never uploaded and thus inaccessible to/safe from the web client). For ease of visualization, plots of viral load distributions are shown as kernel-density estimates (i.e. smoothed) using a Gaussian kernel of width 0.25 log10 viral load units (~1.7-fold).

Statistical tests. The geometric mean viral load for each patient group was calculated as a summary statistic. The geometric (as opposed to arithmetic) mean was chosen because viral loads vary over many orders of magnitude.[2] The Kolmogorov-Smirnov test (KS; scipy.stats.kstest) was used to compare distributions. This test was used because data were not distributed normally and KS does not require normality (unlike, for example, the t-test, which requires normal distributions). KS tests the null hypothesis that the distributions of viral loads for two patient groups are statistically indistinguishable.[25] The p-value gives the probability that distributions from the two groups are drawn from the same underlying distribution. A large p-value means the two groups are statistically indistinguishable; a small p-value means they are different. Interpretation of p-values as significant vs. not significant requires a significance threshold, which requires correction for multiple comparisons if multiple comparisons are performed.[26,27] Because the number of comparisons performed via the web portal is up to the user, un-corrected p-values are reported, with interpretation as significant or not significant left to the user.

Software and hardware. Data extraction, annotation, statistics, and analyses were performed using standard Unix tools and Python 3.9+ using the pandas, numpy, scipy, and scikit libraries and the interactive Jupyter notebook environment. Figures were created using Python graphics libraries matplotlib and seaborn, and OmniGraffle 7 (The Omni Group, Seattle, WA).

Results

A web portal for large-scale real-world SARS-CoV-2 viral load results for different patient groups. 46,726 COVID-19 PCR results representing approximately 39,180 unique individuals were converted to viral loads and annotated for patient demographics, comorbidities, presentation, treatment, and socioeconomic status and made available for interactive investigation via a public web portal at https://arnaoutlab.org/coviral/ (Table 1). The portal[28,29] allows users to visualize the viral load distribution for any patient group, to compare distributions between groups, and to estimate, for each group, the sensitivity and specificity of a given OTC test for detecting contagious individuals. Users can define and compare complex subgroups by selecting multiple characteristics via checkboxes in the user interface (Fig. 1). In this work, all the figures that contain distributions are direct screenshots from the portal.

Overall viral load distributions. Viral loads varied over nearly 10 orders of magnitude, from 7 copies/mL (the lowest our system will report) to 1.5 billion copies/mL (99th percentile). This extraordinary range is consistent with observations from early in the pandemic (spring-summer of 2020).[2] The referenced early observations suggested that viral loads were, to a good approximation, distributed fairly uniformly over the range. In contrast, the current dataset, which is ten times as large (46,726 results vs. 4,774 in the previous work[2]), demonstrates clear bimodality: patients' viral loads tend to be either very low, with a peak around the LOD of 100 copies/mL, or else very high, with a peak around 100 million copies/mL (Fig. 2). This bimodality is apparent in retrospect (e.g., Fig. 2a of Arnaout et al. 2021[2]) but required a large dataset to visualize clearly. Further research is needed to understand the reason(s) for these two peaks.

Viral load comparisons among patient groups: remdesivir treatment and patient presentation. The web portal allows statistical comparisons of many thousands of specific patient groups. Here we describe several examples that are illustrative of the questions that can be asked and answered using this resource. Remdesivir (Gilead Sciences, Foster City, CA) is an intravenously administered RNA polymerase inhibitor initially approved by the FDA for treatment of SARS-CoV-2 in hospitalized adults and adolescents.[30] Of the 46,726 test results in our dataset, 688 were from patients who then received remdesivir treatment. In practice, at our institution, remdesivir was used for sicker patients. We hypothesized that viral loads would be higher on average in patients who received remdesivir and in sicker patients. The data supported this hypothesis: viral loads were higher on average in both remdesivir-receiving and sicker-appearing patients, with means of 8.6×10^5 copies/mL in patients who received remdesivir vs. 1.2×10^5 in those who did not (Fig. 3a) and 9.4×10^4 in sick-appearing patients vs. 4.1×10^4 in well-appearing patients (Fig. 3b). In both cases, the difference was due to a greater fraction patients in the high-viral-load peak. This was especially clear in the remdesivir comparison (Fig. 3a). In each case, the KS p-value was 4.3×10^{-16}, which we interpret as rejecting the null hypothesis of no difference, with high confidence. These are examples in which the data confirmed hypotheses regarding differences in viral load.

Unexpected findings: pulmonary disease. Serious cases of COVID-19 are marked by life-threatening respiratory distress. This evolution became less common with the emergence of the omicron strain and the increasing prevalence of prior immunological exposure including vaccination. We hypothesized that patients with pulmonary disease would have higher viral loads than patients without pulmonary disease, especially for early viral variants, which had a stronger tropism for lung as opposed to the upper respiratory tract. However, this hypothesis was not supported (Fig. 3c). Viral loads for the 975 patients with pulmonary disease tested during the early-variant era were statistically indistinguishable from those for the 27,308 patients with no pulmonary disease who were tested during the same time period. This is an example of unexpected findings that this dataset and its web-portal interface can reveal.

Comparisons among multiple groups: survivorship and causes of death. The web portal also allows users to compare more than two groups of patients at a time. In quantifying mortality during the pandemic, one distinction of value has been between individuals who died with COVID-19 as the proximal cause of death and individuals who died with COVID-19 as an incidental finding. We compared these two groups with survivors (Fig. 3d). We found that the 398 patients who died from COVID-19 in our dataset had higher viral loads than either of the other two groups, and that viral loads were statistically indistinguishable between the approximately 46,000 survivors and the 143 patients who died with COVID-19 as an incidental finding (p=0.07). For ease of comparison, the web portal displays distributions in a ridgeline plot from lowest to highest mean, top to bottom. When there are three or more groups, p-values are displayed as a heatmap, accompanied by explanatory text. (Because KS p-values are symmetric, only the top half of the heatmap is shown.)

Complex patient subgroups: race and presumed variant. The ability to interrogate complex subgroups by checking multiple boxes in the web-portal interface allows more subtle investigations. For example, Black patients have experienced disproportionate morbidity and mortality during the pandemic.[31] However, the 13,806 patients who self-reported as White in our dataset on average had slightly higher viral loads than the 8,299 who self-reported as Black (KS $p=3.3\times10^{-13}$). That the viral loads in the White group were on average *higher* suggests that differences in outcome between these groups are not explained by differences in viral load (Fig. 3f-g), despite the clear relationship between viral load and survivorship described above (Fig. 3d). Interestingly, the observed difference is more pronounced during the delta-variant wave (Fig. 3h). During the delta wave (July 7, 2021 to December 6, 2021), viral loads were on average three times as high for White patients (8.0×10^5 copies/mL, n=1,024) as Black patients (2.7×10^5 copies/mL, n=665; KS test $p=5.2\times10^{-8}$) with a distinctly sharper high-viral-load peak in White patients. This difference was greater in patients over 30 years old and was almost entirely absent in patients under 30 (30-60 y.o: 398 Black patients and 719 White patients, $p= 2.5\times10^{-5}$; <30 y.o.: 266 Black vs. 299 White patients, p=0.14). In contrast, viral load distributions for Black and White patients were more similar for both early in the pandemic and during the omicron variant time period ($p=4.0\times10^{-5}$ for 4,393 Black and 7,042 White patients and p=0.02 for 2,295

Black and 4,341 White patients, respectively). This example illustrates the utility and (statistical) power of the portal for investigating subgroups of interest.

Antigen test performance. In the head-to-head comparison of PCR and antigen test results, 281 patients consented to participate. Of the PCR samples collected, 277 were tested; the remaining four were mishandled or leaked. Of the 277, 65 had a positive COVID-19 result by PCR (23%). The PCR-positive samples were all tested on either the Alinity m SARS-CoV-2 real time RT-PCR assay or the Alinity m Resp-4-Plex PCR assay. Viral loads in the PCR-positive patients ranged from approximately 10 to approximately 10^9 copies/mL, with a peak in the distribution between 10^6 and 10^8. Of the 65 positive samples, 3 were sequenced and 20 selected at random were used to assess contagiousness in viral culture.

Antigen test performance. Of the 65 patients with positive PCR tests, 43 tested positive on the Binax antigen test and 40 tested positive on the CareStart antigen test. No invalid antigen tests (lacking the control line) were observed. Only one of the patients who tested negative by PCR tested positive on the antigen tests (both Binax and CareStart), confirming the high specificity of these tests. The proportion of positive antigen tests varied with viral load. At viral loads less than 10^3 copies/mL, both antigen tests were always negative; at viral loads greater than 10^7 copies/mL, both antigen tests were always positive. However, there was an overlap of antigen-test-positive and antigen-test-negative results at intermediate viral loads (Fig. 4a). k and v_0 values (see Methods) were comparable between the two antigen tests (k=1.184, v_0=4.538 for Binax and k=1.142, v_0=4.995 for CareStart). The resulting S-shaped curves were used to predict antigen test performance in the web portal.

The OTC antigen tests that have been widely available on the market since 2021 are considerably less sensitive than RT-qPCR for detecting SARS-CoV-2 infection. However, because their LODs are generally above the contagiousness threshold, they are quite sensitive for detecting contagiousness.[10] Based on our clinical experience, we hypothesized that antigen tests would perform similarly on different patient groups and subgroups; this hypothesis was largely supported (Fig. 4b-d). The web portal also allows users to estimate sensitivity and

specificity for the BinaxNow COVID-19 Ag Card and CareStart COVID-19 Antigen Home Test, based on the modelled performance curves, on any sufficiently large user-selected patient group (Fig 6). The two tests performed well: sensitivities for detecting contagiousness were roughly 0.85-0.90 across patient groups.

Contagiousness for omicron-era virus. For early-pandemic and delta-wave strains of SARS-CoV-2, the threshold viral load for contagiousness has previously been found to be approximately 10^5 copies/mL.[10] For omicron variants, we found that the threshold is statistically indistinguishable from this, at 4.5×10^4 copies/mL (confidence interval, 1.1×10^4-1.9×10^5 copies/mL; p=0.23, Fig. 5). The omicron threshold was based on 20 PCR-positive results from our head-to-head clinical trial of 277 total patients. We confirmed that the dominant strain circulating in the Massachusetts Bay area was omicron (BA5.2/Clade 22B) via next-generation sequencing, with only rare single-nucleotide differences relative to those already described (Fig. 6a-d). Because the strain mix in Massachusetts (Fig. 6e) has been highly representative of the strain mix in the country as a whole throughout the pandemic (Fig. 6f), these results support the generalizability of these findings from a particular geographic area to the entire population.

Discussion

The COVID-19 pandemic has proven a catalyst for accelerating medical advances, including the development of more efficient methods for developing and testing critical diagnostic assays.[32–37] It has also drawn attention to the value of reliable public data, including large public datasets.[38–42] Here we describe such a dataset, to our knowledge the first large dataset of SARS-CoV-2 viral loads in patients across the history of the pandemic through to the present day. The rich clinical annotations of this dataset reveal similarities and differences in viral loads among patients by demographics, presentation, and comorbidity as well as by vaccination status, treatment, and socioeconomic status. Protected patient data was safeguarded by multiple mechanisms. The data revealed cases in which differences were expected as well as cases in which they were unexpected. The size of the dataset and extent of the annotation allow more comparisons than can reasonably be summarized in a single publication; availability of this

dataset via the web allows anyone—clinicians, investigators, developers, regulators, and patients, alone or with the assistance of artificial intelligence and/or machine learning (AI/ML)-based tools—to explore and conduct research, to test existing hypotheses, and to generate new research questions.

The clinical utility of measuring and investigating viral load in SARS-CoV-2 has been demonstrated.[43–46] This is consistent with both the advantage of viral loads relative to Ct values[47,48] and their utility in earlier viral infections such as HIV and hepatitis C (HCV). SARS-CoV-2 viral loads have already proven useful in the development and characterization of COVID-19 diagnostics in multiple contexts, including testing on nasopharyngeal secretions, nasal secretions, and saliva.[5,6] Here we demonstrate their potential to regulators as a tool to streamline evaluation of new OTC tests. Specifically, we demonstrate that a test's analytical performance measure, namely the LOD, can be used to estimate that test's clinical sensitivity for detecting contagiousness in any patient group, without having to conduct a dedicated clinical trial for that group. The alternative of innumerable dedicated trials is beyond reasonable expectations of the financial capability of developers or the bandwidth of their clinical testing partners. For this reason regulatory agencies such as the FDA have expressed interest in methods that use large-scale real-world data to streamline test evaluation.

Two assumptions implicit in this approach are worth mention. First, we model the success rate of antigen tests solely as a function of viral load. The assumption is that no other non-negligible factor varies systematically between patient groups. Second, we assume that the curve—success rate as a function of viral load—can be adequately predicted from the data from a fairly small study. There are two possible sources of error in this curve: sampling error, which can be reduced by increasing the number of subjects sampled; and lack-of-fit error, the error inherent in trying to fit a function of the wrong form. The function used in the web portal's calculations, was chosen on the basis of its history of use in dose-response-type situations, but carries the (reasonable) assumptions that the probability increases smoothly and continuously with increasing viral load, and approaches 0 with sufficiently low viral load and 1 with sufficiently high viral load. These constraints leave only two free parameters, which is desirable for

statistical power and robustness to the noise inherent in any such study (e.g., sampling error and measurement error).

A user of the web portal who is accustomed to thinking of test quality solely in terms of LOD (limit of detection) might initially be surprised that the portal prefers two parameters, not one, to define an antigen test's performance. In effect, the LOD parameter sets the location of the S-shaped curve that relates viral load and performance, and the second parameter—here, the 50% detection threshold—sets the S-shaped curve's steepness. Without the second parameter, one could fit a curve in which the sensitivity is 0% at all viral loads below the LoD, and 95% at any higher viral load, which is clearly quite different from the relationship observed in the head-to-head study. How different the true shape of the antigen test performance curve is from the logit function used here, and thus whether fitting a different function would work better, can be elucidated by larger head-to-head studies; however, in the midst of a public health emergency, the cost in time of sampling more subjects must be weighed against the value of complete certainty. At any rate, the fitting error was low.

Three important limitations to the dataset in its current form also deserve mention. The first is due to incompleteness of some of the data fields, for example presentation and vaccination status (Table 1). Presentation information was only sometimes available in structured form in our data repository; we did not attempt to extract data from notes to complement incomplete records. Vaccination status was likewise only sometimes available in a structured manner; integration with state-level records could potentially fill in missing records. Second, patient-level annotations are not yet available for download as part of the dataset. Making viral loads freely and easily available for patient groups required significant attention to avoid potential loopholes that might risk patient privacy via identifiability. Methods included suppressing data transmission for groups that were small enough to present potential "journalist risk,"[49] jittering counts to prevent deduction of the sizes of suppressed groups, and rounding viral loads to two log-scale decimal places.[50] Further work is necessary to make patient-level annotations available. Third and finally, the size of the dataset, while large, is still insufficient for the smallest groups—for example, cystic fibrosis patients, patients who recently delivered, or

Native Americans—to be sufficiently sizeable to draw robust conclusions. One solution is to add data from other care-giving institutions that performed substantial COVID-19 testing; another is to supplement existing large datasets, for example the 50-million-person CVD-COVID-UK initiative, with viral loads. The free availability of methods to convert from Ct values to viral loads facilitates such advances.[4]

As part of the drive toward precision medicine, clinical care benefits from personalization of diagnostic testing: the right test for the right patient, where the importance of patient heterogeneity is increasingly accepted. We have demonstrated that large-scale real-world data can assist this effort by enabling the estimation of personalized clinical sensitivities and specificities without the need for dedicated clinical trials on every patient group. This work is generalizable beyond COVID-19. Laboratory testing is an exceptionally rich source of real-world medical information. It is the highest volume medical activity, with some 20 billion tests performed each year in the United States alone. It is also the most cost-effective, costing just pennies on the healthcare dollar. It is integral to decision-making across medicine, for patients at every level of acuity, from screening to emergencies. Its results are almost always numerical or categorical, making it especially amenable to modern approaches like machine learning. And computational re-analysis is substantially less expensive than de novo trials. The present work supports the view that meaningful value can come economically from additional repurposing of the vast stores of real-world laboratory results that exist in healthcare institutions.

Acknowledgements

The authors acknowledge Elliot Hill for assistance on plugging ct2vl into the data processing workflow; to Timothy Graham, Gail Piatkowski, Baevin Feeser, and Griffin Weber for their advice and guidance on extracting data from EMR databases; to funding from the Reagan-Udall Foundation for the FDA (R.A.); and to funding via a Novel Therapeutics Delivery Grant from the Massachusetts Life Sciences Center (J.E.K.). We would also like to acknowledge Abbott (Scarborough, ME) for provision of the Binax Now antigen tests and Ginkgo Bioworks (Boston, MA) for provision of CareStart antigen tests used in this study.

Author Contributions

Conceptualization, J.K. and R.A.; Software, A.M.; Investigation, D.H., E.C., S.D., and M.Y.; Resources: S.R.; Data Curation, A.M.; Writing — Original Draft, R.A. and A.M.; Writing — Review & Editing, P.K. and J.K.; Visualization, T.S., D.B., A.M., and R.A.; Supervision, R.A; Funding Acquisition, P.K., S.R., and R.A.

Declaration of interests

The authors declare no competing interests.

Table and Figure Legends

Table 1: Summary of patient characteristics. Shown are counts for select high-level categories as of April 14, 2023. Counts may differ somewhat from the counts presented through the web portal as a result of jittering and as more data continues to be added through the portal. Note that counts broken down by characteristics do not add up to the total, because of the nulling out of some data to reduce re-identification risk (see Methods).

Table S1: Detailed criteria for defining patient characteristics, Related to Methods. Criteria for each of the characteristics available on the portal are described.

Fig. 1: Patient characteristics available for defining patient groups and subgroups. This screenshot from the web portal demonstrates the ability for users to define patient groups by demographic, comorbidity, treatment, vaccination status, pandemic era, and so on, using checkboxes.

Fig. 2: Overall bimodal distribution of viral loads. When no checkboxes are selected to constrain or partition the dataset, users see the distribution of all viral loads. The marked bimodal distribution is clearly apparent.

Fig. 3: Viral load comparisons by remdesivir treatment, patient presentation, and outcome. Screenshot from the web portal. Viral loads are higher in (a) patients who received remdesivir vs. patients who did not receive remdesivir and in (b) patients who presented as ill-appearing vs. patients who presented as well-appearing (see Table S1 for precise definitions). Note the bimodal distributions, with a low-viral-load peak and a high-viral-load peak. (c) Viral load distributions and pairwise p-values by presence or absence of pulmonary disease for patients during the early variant era. (d) Patients who died from COVID-19 had higher viral loads than either patients who died with COVID-19 as an incidental finding or survivors. Viral loads for were statistically indistinguishable between the latter two groups. The web portal displays distributions in a ridgeline plot from lowest to highest mean, top to bottom. (e) Because there are three or more groups, p-values are displayed as a heatmap, accompanied by explanatory text. Because KS p-values are symmetric, only the top half of the heatmap is shown. Fig. 3f-h: viral load distributions by self-reported race and by race-plus-presumed variant. (f) Viral load distributions by race. (g) Statistical comparisons among these distributions. (h) Viral load distributions between Black vs. White patients in the delta-variant era.

Fig. 4: Antigen test results from head-to-head trial and performance on patient subgroups. (a) Each antigen test result for each PCR-positive patient, vs. $\log_{10}$ of viral load according to the simultaneous PCR test. (b) Antigen test performance on patients in neighborhoods stratified by median household income. (c) Performance of BinaxNow COVID-19 Ag Card on patients with median household income >$130,000. (d) Performance of BinaxNow COVID-19 Ag Card on patients with median household income <$52,000. The user can use the radio buttons to select any of the patient groups in the viral load distributions in the top section of the web portal. The imputed positive antigen-test results will appear as a lighter shade, and the negative results a darker shade.

Fig. 5: Determination of the contagiousness threshold for omicron-era SARS-CoV-2 strains. Below a certain viral load in the patient sample (x-axis), no virus was recoverable from culture (y-axis, maximum of day 3 and day 6 supernatants). For viewing convenience, culture-negative samples are plotted at 0.1 copies/mL (dotted line, 1 copy/mL). The gray region shows the

confidence interval for the threshold. The red line shows the midpoint of this region (on the log10 scale), 45,000 copies/mL, as the most likely threshold.

Fig. 6. Sequencing late-2022 strain and generalizability of Massachusetts-level results to the United States as a whole. Results of sequencing of a BA5.2/Clade 22B patient sample from Aug.-Sep. 2022 (97.6% coverage). (a) Sample relative to COVID-19 phylogeny (with clade labels). (b) First 64 of the 72 nucleotide substitutions relative to the original Wuhan strain. (c) 52 amino acid substitutions relative to the Wuhan strain. (d) The five unique ("private") mutations relative to the phylogenetic tree. (e) Distribution of strains in Massachusetts near the time of the sample according to covariants.org. (f) Comparison by frequency of the strains circulating in Massachusetts to those circulating in the United States at the same times demonstrating generalizability of Massachusetts-state variant patterns to the country as a whole. Red line, 1:1. Gray, early strains; purple, delta strains; green, omicron strains. R^2 is for least-squares linear regression of USA vs. Massachusetts data (regression slope=0.97, intercept=0.00).

Tables and Figures

Table 1: Summary of patient characteristics

Sex	
Female	25,884
Male	20,608
Age	
<30 years old	12,446
30-60 years old	21,889
> 60 years old	12,100
Self-reported Race or Ethnicity	
Unknown/Other	14,530
White	13,806
Black	8,299
Hispanic	7,540
Asian/Pacific Islander	2,472
Setting	
Inpatient	2,157
Outpatient	11,758
Emergency room	1,779
Other institutions	31,031
Variant	
Early	28,289
Delta	2,911
Omicron	11,264
Vaccination status	
Vaccinated	6,806
Unvaccinated	6,960
Unknown	32,732
Outcome	
Died from COVID-19	398
Died with COVID-19 as an incidental finding	143
Survived	45,938
Testing platform	
Abbott m2000	24,243
Abbott Alinity	20,593
Abbott Alinity 4-plex	1,889
Total	
	46,726

Table S1: Detailed criteria for defining patient characteristics, Related to Methods

Table 1

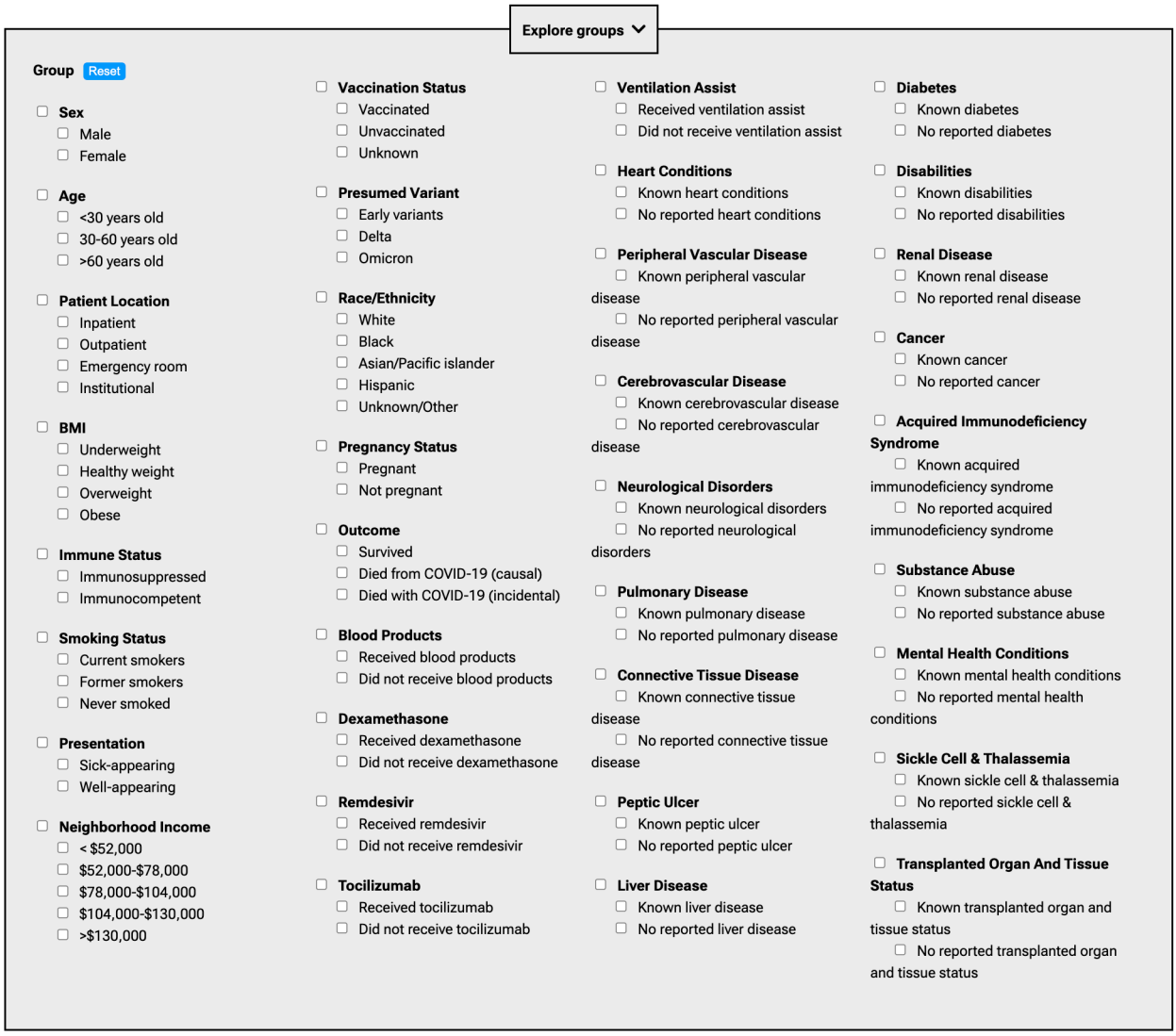
Checkbox name	Notes
Acquired immunodeficiency syndrome	ICD10 codes: ["B20","O98711","O98712","O98713","O98719","O9872","O9873","Z21"]
Age	years
Blood products	CPT4 codes: ["P9012","P9016","P9047","P9059","P9073"]
BMI	kilograms/(meter**2); allowed values: {"Underweight": "bmi < 18.5 AND age > 17", "Healthy weight": "bmi >= 18.5 AND bmi < 25 AND age > 17", "Overweight": "bmi >= 25 AND bmi < 30 AND age > 17", "Obese": "bmi >= 30 AND age > 17"}
Cancer	ICD10 codes: ['C000', 'C001', 'C002', 'C003', 'C004', 'C005', 'C006', 'C008', 'C009', 'C01', 'C020', 'C021', 'C022', 'C023', 'C024', 'C028', 'C029', 'C030', 'C031', 'C039', 'C040', 'C041', 'C048', 'C049', 'C050', 'C051', 'C052', 'C058', 'C059', 'C060', 'C061', 'C062', 'C0680', 'C0689', 'C069', 'C07', 'C080', 'C081', 'C089', 'C090', 'C091', 'C098', 'C099', 'C100', 'C101', 'C102', 'C103', 'C104', 'C108', 'C109', 'C110', 'C111', 'C112', 'C113', 'C118', 'C119', 'C12', 'C130', 'C131', 'C132', 'C138', 'C139', 'C140', 'C142', 'C148', 'C153', 'C154', 'C155', 'C158', 'C159', 'C160', 'C161', 'C162', 'C163', 'C164', 'C165', 'C166', 'C168', 'C169', 'C170', 'C171', 'C172', 'C173', 'C178', 'C179', 'C180', 'C181', 'C182', 'C183', 'C184', 'C185', 'C186', 'C187', 'C188', 'C189', 'C19', 'C20', 'C210', 'C211', 'C212', 'C218', 'C220', 'C221', 'C222', 'C223', 'C224', 'C227', 'C228', 'C229', 'C23', 'C240', 'C241', 'C248', 'C249', 'C250', 'C251', 'C252', 'C253', 'C254', 'C257', 'C258', 'C259', 'C260', 'C261', 'C269', 'C300', 'C301', 'C310', 'C311', 'C312', 'C313', 'C318', 'C319', 'C320', 'C321', 'C322', 'C323', 'C328', 'C329', 'C33', 'C3400', 'C3401', 'C3402', 'C3410', 'C3411', 'C3412', 'C342', 'C3430', 'C3431', 'C3432', 'C3480', 'C3481', 'C3482', 'C3490', 'C3491', 'C3492', 'C37', 'C380', 'C381', 'C382', 'C383', 'C384', 'C388', 'C390', 'C399', 'C4000', 'C4001', 'C4002', 'C4010', 'C4011', 'C4012', 'C4020', 'C4021', 'C4022', 'C4030', 'C4031', 'C4032', 'C4080', 'C4081', 'C4082', 'C4090', 'C4091', 'C4092', 'C410', 'C411', 'C412', 'C413', 'C414', 'C419', 'C430', 'C4310', 'C4311', 'C43111', 'C43112', 'C4312', 'C43121', 'C43122', 'C4320', 'C4321', 'C4322', 'C4330', 'C4331', 'C4339', 'C434', 'C4351', 'C4352', 'C4359', 'C4360', 'C4361', 'C4362', 'C4370', 'C4371', 'C4372', 'C438', 'C439', 'C4400', 'C4409', 'C44101', 'C44102', 'C441021', 'C441022', 'C44109', 'C441091', 'C441092', 'C44131', 'C441321', 'C441322', 'C441391', 'C441392', 'C44191', 'C44192', 'C441921', 'C441922', 'C44199', 'C441991', 'C441992', 'C44201', 'C44202', 'C44209', 'C44291', 'C44292', 'C44299', 'C44300', 'C44301', 'C44309', 'C44390', 'C44391', 'C44399', 'C4440', 'C4449', 'C44500', 'C44501', 'C44509', 'C44590', 'C44591', 'C44599', 'C44601', 'C44602', 'C44609', 'C44691', 'C44692', 'C44699', 'C44701', 'C44702', 'C44709', 'C44791', 'C44792', 'C44799', 'C4480', 'C4489', 'C4490', 'C4499', 'C450', 'C451', 'C457', 'C459', 'C460', 'C461', 'C462', 'C463', 'C464', 'C4650', 'C4651', 'C4652', 'C467', 'C469', 'C470', 'C4710', 'C4711', 'C4712', 'C4720', 'C4721', 'C4722', 'C473', 'C474', 'C475', 'C476', 'C478', 'C479', 'C480', 'C481', 'C482', 'C488', 'C490', 'C4910', 'C4911', 'C4912', 'C4920', 'C4921', 'C4922', 'C493', 'C494', 'C495', 'C496', 'C498', 'C499', 'C49A0', 'C49A1', 'C49A2', 'C49A3', 'C49A4', 'C49A5', 'C49A9', 'C4A0', 'C4A10', 'C4A11', 'C4A111', 'C4A112', 'C4A12', 'C4A121', 'C4A122', 'C4A20', 'C4A21', 'C4A22', 'C4A30', 'C4A31', 'C4A39', 'C4A4', 'C4A51', 'C4A52', 'C4A59', 'C4A60', 'C4A61', 'C4A62', 'C4A70', 'C4A71', 'C4A72', 'C4A8', 'C4A9', 'C50011', 'C50012', 'C50019', 'C50021', 'C50022', 'C50029', 'C50111', 'C50112', 'C50119', 'C50121', 'C50122', 'C50129', 'C50211', 'C50212', 'C50219', 'C50221', 'C50222', 'C50229', 'C50311', 'C50312', 'C50319', 'C50321', 'C50322', 'C50329', 'C50411', 'C50412', 'C50419', 'C50421', 'C50422', 'C50429', 'C50511', 'C50512', 'C50519', 'C50521', 'C50522', 'C50529', 'C50611', 'C50612', 'C50619', 'C50621', 'C50622', 'C50629', 'C50811', 'C50812', 'C50819', 'C50821', 'C50822', 'C50829', 'C50911', 'C50912', 'C50919', 'C50921', 'C50922', 'C50929', 'C510', 'C511', 'C512', 'C518', 'C519', 'C52', 'C530', 'C531', 'C538', 'C539', 'C540', 'C541', 'C542', 'C543', 'C548', 'C549', 'C55', 'C561', 'C562', 'C563', 'C569', 'C5700', 'C5701', 'C5702', 'C5710', 'C5711', 'C5712', 'C5720', 'C5721', 'C5722', 'C573', 'C574', 'C577', 'C578', 'C579', 'C58', 'C600', 'C601', 'C602', 'C608', 'C609', 'C61', 'C6200', 'C6201', 'C6202', 'C6210', 'C6211', 'C6212', 'C6290', 'C6291', 'C6292', 'C6300', 'C6301', 'C6302', 'C6310', 'C6311', 'C6312', 'C632', 'C637', 'C638', 'C639', 'C641', 'C642', 'C649', 'C651', 'C652', 'C659', 'C661', 'C662', 'C669', 'C670', 'C671', 'C672', 'C673', 'C674', 'C675', 'C676', 'C677', 'C678', 'C679', 'C680', 'C681', 'C688', 'C689', 'C6900', 'C6901', 'C6902', 'C6910', 'C6911', 'C6912', 'C6920', 'C6921', 'C6922', 'C6930', 'C6931', 'C6932', 'C6940', 'C6941', 'C6942', 'C6950', 'C6951', 'C6952', 'C6960', 'C6961', 'C6962', 'C6980', 'C6981', 'C6982', 'C6990', 'C6991', 'C6992', 'C700', 'C701', 'C709', 'C710', 'C711', 'C712', 'C713', 'C714', 'C715', 'C716', 'C717', 'C718', 'C719', 'C720', 'C721', 'C7220', 'C7221', 'C7222', 'C7230', 'C7231', 'C7232', 'C7240', 'C7241', 'C7242', 'C7250', 'C7259', 'C729', 'C73', 'C7400', 'C7401', 'C7402', 'C7410', 'C7411', 'C7412', 'C7490', 'C7491', 'C7492', 'C750', 'C751', 'C752', 'C753', 'C754', 'C755', 'C758', 'C759', 'C760', 'C761', 'C762', 'C763', 'C7640', 'C7641', 'C7642', 'C7650', 'C7651', 'C7652', 'C768', 'C770', 'C771', 'C772', 'C773', 'C774', 'C775', 'C778', 'C779', 'C7800', 'C7801', 'C7802', 'C781', 'C782', 'C7830', 'C7839', 'C784', 'C785', 'C786', 'C787', 'C7880', 'C7889', 'C7900', 'C7901', 'C7902', 'C7910', 'C7911', 'C7919', 'C792', 'C7931', 'C7932', 'C7940', 'C7949', 'C7951', 'C7952', 'C7960', 'C7961', 'C7962', 'C7963', 'C7970', 'C7971', 'C7972', 'C7981', 'C7982', 'C7989', 'C799', 'C7A00', 'C7A010', 'C7A011', 'C7A012', 'C7A019', 'C7A020', 'C7A021', 'C7A022', 'C7A023', 'C7A024', 'C7A025', 'C7A026', 'C7A029', 'C7A090', 'C7A091', 'C7A092', 'C7A093', 'C7A094', 'C7A095', 'C7A096', 'C7A098', 'C7A1', 'C7A8', 'C7B00', 'C7B01', 'C7B02', 'C7B03', 'C7B04', 'C7B09', 'C7B1', 'C7B8', 'C800', 'C8100', 'C8101', 'C8102', 'C8103', 'C8104', 'C8105', 'C8106', 'C8107', 'C8108', 'C8109', 'C8110', 'C8111', 'C8112', 'C8113', 'C8114', 'C8115', 'C8116', 'C8117', 'C8118', 'C8119', 'C8120', 'C8121', 'C8122', 'C8123', 'C8124', 'C8125', 'C8126', 'C8127', 'C8128', 'C8129', 'C8130', 'C8131', 'C8132', 'C8133', 'C8134', 'C8135', 'C8136', 'C8137', 'C8138', 'C8139', 'C8140', 'C8141', 'C8142', 'C8143', 'C8144', 'C8145', 'C8146', 'C8147', 'C8148', 'C8149', 'C8170', 'C8171', 'C8172', 'C8173', 'C8174', 'C8175', 'C8176', 'C8177', 'C8178', 'C8179', 'C8190', 'C8191', 'C8192', 'C8193', 'C8194', 'C8195', 'C8196', 'C8197', 'C8198', 'C8199', 'C8200', 'C8201', 'C8202', 'C8203', 'C8204', 'C8205', 'C8206', 'C8207', 'C8208', 'C8209', 'C8210', 'C8211', 'C8212', 'C8213', 'C8214', 'C8215', 'C8216', 'C8217', 'C8218', 'C8219', 'C8220', 'C8221', 'C8222', 'C8223', 'C8224', 'C8225', 'C8226', 'C8227', 'C8228', 'C8229', 'C8230', 'C8231', 'C8232', 'C8233', 'C8234', 'C8235', 'C8236', 'C8237', 'C8238', 'C8239', 'C8240', 'C8241', 'C8242', 'C8243', 'C8244', 'C8245', 'C8246', 'C8247', 'C8248', 'C8249', 'C8250', 'C8251', 'C8252', 'C8253', 'C8254', 'C8255', 'C8256', 'C8257', 'C8258', 'C8259', 'C8260', 'C8261', 'C8262', 'C8263', 'C8264', 'C8265', 'C8266', 'C8267', 'C8268', 'C8269', 'C8280', 'C8281', 'C8282', 'C8283', 'C8284', 'C8285', 'C8286', 'C8287', 'C8288', 'C8289', 'C8290', 'C8291', 'C8292', 'C8293', 'C8294', 'C8295', 'C8296', 'C8297', 'C8298', 'C8299', 'C8300', 'C8301', 'C8302', 'C8303', 'C8304', 'C8305', 'C8306', 'C8307', 'C8308', 'C8309', 'C8310', 'C8311', 'C8312', 'C8313', 'C8314', 'C8315', 'C8316', 'C8317', 'C8318', 'C8319', 'C8330', 'C8331', 'C8332', 'C8333', 'C8334', 'C8335', 'C8336', 'C8337', 'C8338', 'C8339', 'C8350', 'C8351', 'C8352', 'C8353', 'C8354', 'C8355', 'C8356', 'C8357', 'C8358', 'C8359', 'C8370', 'C8371', 'C8372', 'C8373', 'C8374', 'C8375', 'C8376', 'C8377', 'C8378', 'C8379', 'C8380', 'C8381', 'C8382', 'C8383', 'C8384', 'C8385', 'C8386', 'C8387', 'C8388', 'C8389', 'C8390', 'C8391', 'C8392', 'C8393', 'C8394', 'C8395', 'C8396', 'C8397', 'C8398', 'C8399', 'C8400', 'C8401', 'C8402', 'C8403', 'C8404', 'C8405', 'C8406', 'C8407', 'C8408', 'C8409', 'C8410', 'C8411', 'C8412', 'C8413', 'C8414', 'C8415', 'C8416', 'C8417', 'C8418', 'C8419', 'C8440', 'C8441', 'C8442', 'C8443', 'C8444', 'C8445', 'C8446', 'C8447', 'C8448', 'C8449', 'C8460', 'C8461', 'C8462', 'C8463', 'C8464', 'C8465', 'C8466', 'C8467', 'C8468', 'C8469', 'C8470', 'C8471', 'C8472', 'C8473', 'C8474', 'C8475', 'C8476', 'C8477', 'C8478', 'C8479', 'C847A', 'C8490', 'C8491', 'C8492', 'C8493', 'C8494', 'C8495', 'C8496', 'C8497', 'C8498', 'C8499', 'C84A0', 'C84A1', 'C84A2', 'C84A3', 'C84A4', 'C84A5', 'C84A6', 'C84A7', 'C84A8', 'C84A9', 'C84Z0', 'C84Z1', 'C84Z2', 'C84Z3', 'C84Z4', 'C84Z5', 'C84Z6', 'C84Z7', 'C84Z8', 'C84Z9', 'C8510', 'C8511', 'C8512', 'C8513', 'C8514', 'C8515', 'C8516', 'C8517', 'C8518', 'C8519', 'C8520', 'C8521', 'C8522', 'C8523', 'C8524', 'C8525', 'C8526', 'C8527', 'C8528', 'C8529', 'C8580', 'C8581', 'C8582', 'C8583', 'C8584', 'C8585', 'C8586', 'C8587', 'C8588', 'C8589', 'C8590', 'C8591', 'C8592', 'C8593', 'C8594', 'C8595', 'C8596', 'C8597', 'C8598', 'C8599', 'C860', 'C861', 'C862', 'C863', 'C864', 'C865', 'C866', 'C880', 'C882', 'C883', 'C884', 'C888', 'C889', 'C9000', 'C9001', 'C9002', 'C9010', 'C9011', 'C9012', 'C9020', 'C9021', 'C9022', 'C9030', 'C9031', 'C9032', 'C9100', 'C9101', 'C9102', 'C9110', 'C9111', 'C9112', 'C9130', 'C9131', 'C9132', 'C9140', 'C9141', 'C9142', 'C9150', 'C9151', 'C9152', 'C9160', 'C9161', 'C9162', 'C9190', 'C9191', 'C9192', 'C91A0', 'C91A1', 'C91A2', 'C91Z0', 'C91Z1', 'C91Z2', 'C9200', 'C9201', 'C9202', 'C9210', 'C9211', 'C9212', 'C9220', 'C9221', 'C9222', 'C9230', 'C9231', 'C9232', 'C9240', 'C9241', 'C9242', 'C9250', 'C9251', 'C9252', 'C9260', 'C9261', 'C9262', 'C9290', 'C9291', 'C9292', 'C92A0', 'C92A1', 'C92A2', 'C92Z0', 'C92Z1', 'C92Z2', 'C9300', 'C9301', 'C9302', 'C9310', 'C9311', 'C9312', 'C9330', 'C9331', 'C9332', 'C9390', 'C9391', 'C9392', 'C92Z0', 'C93Z1', 'C93Z2', 'C9400', 'C9401', 'C9402', 'C9420', 'C9421', 'C9422', 'C9430', 'C9431', 'C9432', 'C9440', 'C9441', 'C9442', 'C946', 'C9480', 'C9481', 'C9482', 'C9500', 'C9501', 'C9502', 'C9510', 'C9511', 'C9512', 'C9590', 'C9591', 'C9592', 'C960', 'C962', 'C9620', 'C9621', 'C9622', 'C9629', 'C964', 'C969', 'C96A', 'C962', 'D469', 'D472', 'E3121', 'E3122', 'E3123']
Cerebrovascular disease	ICD10 codes: ['G450', 'G451', 'G452', 'G453', 'G454', 'G458', 'G459', 'G460', 'G461', 'G462', 'G463', 'G464', 'G465', 'G466', 'G467', 'G468', 'H3400', 'H3401', 'H3402', 'H3403', 'H3410', 'H3411', 'H3412', 'H3413', 'H34211', 'H34212', 'H34213', 'H34219', 'H34231', 'H34232', 'H34233', 'H34239', 'I6000', 'I6001', 'I6002', 'I6010', 'I6011', 'I6012', 'I602', 'I6020', 'I6021', 'I6022', 'I6030', 'I6031', 'I6032', 'I604', 'I6050', 'I6051', 'I6052', 'I606', 'I607', 'I608', 'I609', 'I610', 'I611', 'I612', 'I613', 'I614', 'I615', 'I616', 'I618', 'I619', 'I6200', 'I6203', 'I621', 'I629', 'I6300', 'I6301', 'I6302', 'I6303', 'I6309', 'I631', 'I6320', 'I6321', 'I6322', 'I6323', 'I6329', 'I6330', 'I6331', 'I6332', 'I6339', 'I634', 'I635', 'I636', 'I638', 'I6381', 'I6389', 'I639', 'I6501', 'I6502', 'I6503', 'I6509', 'I651', 'I6521', 'I6522', 'I6523', 'I6529', 'I658', 'I659', 'I6601', 'I6602', 'I6603', 'I6609', 'I6611', 'I6612', 'I6613', 'I6619', 'I6621', 'I6622', 'I6623', 'I6629', 'I663', 'I668', 'I669', 'I6930', 'I6931', 'I6932', 'I6933', 'I6934', 'I6939', 'I6941', 'I6942', 'I6943', 'I6944', 'I6949', 'I6950', 'I6951', 'I6952', 'I6959', 'I6960', 'I6961', 'I6962', 'I6969', 'I6980', 'I6981', 'I6982', 'I6989', 'I6990', 'I6991', 'I6992', 'I6993', 'I6994', 'I6999', 'I69991', 'I69992', 'I69993', 'I69998', 'P91821', 'P91822', 'P91823', 'P91829']
Connective tissue disease	ICD10 codes: ['M05', 'M06', 'M315', 'M32', 'M33', 'M34', 'M351', 'M353', 'M360']
Presentation	Sick-appearing (any of systolic blood pressure <90 mmHg, diastoloc blood pressure <60 mmHg, heart rate >100 beats per minute, respiratory rate >18 breaths per minute, or temperature >99.1°F within 1 day of PCR test sample) vs. well-appearing
Vaccination status	Vaccinated vs. unvaccinated vs. unknown
Dexamethasone	matches to case-insensitive searches for the string "dexamethasone"
Diabetes	ICD10 codes:['E0800', 'E0801', 'E0810', 'E0811', 'E0821', 'E0822', 'E0829', 'E08311', 'E08319', 'E08321', 'E083211', 'E083212', 'E083213', 'E083219', 'E08329', 'E083291', 'E083292', 'E083293', 'E083299', 'E08331', 'E083311', 'E083312', 'E083313', 'E083319', 'E08339', 'E083391', 'E083392', 'E083393', 'E083399', 'E08341', 'E083411', 'E083412', 'E083413', 'E083419', 'E08349', 'E083491', 'E083492', 'E083493', 'E083499', 'E08351', 'E083511', 'E083512', 'E083513', 'E083519', 'E083521', 'E083522', 'E083523', 'E083529', 'E083531', 'E083532', 'E083533', 'E083539', 'E083541', 'E083542', 'E083543', 'E083549', 'E083551', 'E083552', 'E083553', 'E083559', 'E08359', 'E083591', 'E083592', 'E083593', 'E083599', 'E0836', 'E0837X1', 'E0837X2', 'E0837X3', 'E0837X9', 'E0839', 'E0840', 'E0841', 'E0842', 'E0843', 'E0844', 'E0849', 'E0851', 'E0852', 'E0859', 'E08610', 'E08618', 'E08620', 'E08621', 'E08622', 'E08628', 'E08630', 'E08638', 'E08641', 'E08649', 'E0865', 'E0869', 'E088', 'E089', 'E0900', 'E0901', 'E0910', 'E0911', 'E0921', 'E0922', 'E0929', 'E09311', 'E09319', 'E09321', 'E09322', 'E09323', 'E09329', 'E0936', 'E0937X1', 'E0937X2', 'E0937X3', 'E0937X9', 'E0939', 'E0940', 'E0941', 'E0942', 'E0943', 'E0944', 'E0949', 'E0951', 'E0952', 'E0959', 'E09610', 'E09618', 'E09620', 'E09621', 'E09622', 'E09628', 'E09630', 'E09638', 'E09641', 'E09649', 'E0965', 'E0969', 'E098', 'E099', 'E1010', 'E1011', 'E1021', 'E1022', 'E1029', 'E10311', 'E10319', 'E10321', 'E103211', 'E103212', 'E103213', 'E103219', 'E10329', 'E103291', 'E103292', 'E103293', 'E103299', 'E10331', 'E103311', 'E103312', 'E103313', 'E103319', 'E10339', 'E103391', 'E103392', 'E103393', 'E103399', 'E10341', 'E103411', 'E103412', 'E103413', 'E103419', 'E10349', 'E103491', 'E103492', 'E103493', 'E103499', 'E10351', 'E103511', 'E103512', 'E103513', 'E103519', 'E103521', 'E103522', 'E103523', 'E103529', 'E103531', 'E103532', 'E103533', 'E103539', 'E103541', 'E103542', 'E103543', 'E103549', 'E103551', 'E103552', 'E103553', 'E103559', 'E10359', 'E103591', 'E103592', 'E103593', 'E103599', 'E1036', 'E1037X1', 'E1037X2', 'E1037X3', 'E1037X9', 'E1039', 'E1040', 'E1041', 'E1042', 'E1043', 'E1044', 'E1049', 'E1051', 'E1052', 'E1059', 'E10610', 'E10618', 'E10620', 'E10621', 'E10622', 'E10628', 'E10630', 'E10638', 'E10641', 'E10649', 'E1065', 'E1069', 'E108', 'E109', 'E1100', 'E1101', 'E1110', 'E1111', 'E1121', 'E1122', 'E1129', 'E11311', 'E11319', 'E11321', 'E113211', 'E113212', 'E113213', 'E113219', 'E11329', 'E113291', 'E113292', 'E113293', 'E113299', 'E11331', 'E113311', 'E113312', 'E113313', 'E113319', 'E11339', 'E113391', 'E113392', 'E113393', 'E113399', 'E11341', 'E113411', 'E113412', 'E113413', 'E113419', 'E11349', 'E113491', 'E113492', 'E113493', 'E113499', 'E11351', 'E113511', 'E113512', 'E113513', 'E113519', 'E113521', 'E113522', 'E113523', 'E113529', 'E113531', 'E113532', 'E113533', 'E113539', 'E113541', 'E113542', 'E113543', 'E113549', 'E113551', 'E113552', 'E113553', 'E113559', 'E11359', 'E113591', 'E113592', 'E113593', 'E113599', 'E1136', 'E1137X1', 'E1137X2', 'E1137X3', 'E1137X9', 'E1139', 'E1140', 'E1141', 'E1142', 'E1143', 'E1144', 'E1149', 'E1151', 'E1152', 'E1159', 'E11610', 'E11618', 'E11620', 'E11621', 'E11622', 'E11628', 'E11630', 'E11638', 'E11641', 'E11649', 'E1165', 'E1169', 'E118', 'E119', 'E1300', 'E1301', 'E1310', 'E1311', 'E1321', 'E1322', 'E1329', 'E13311', 'E13319', 'E13321', 'E133211', 'E133212', 'E133213', 'E133219', 'E13329', 'E133291', 'E133292', 'E133293', 'E133299', 'E13331', 'E133311', 'E133312', 'E133313', 'E133319', 'E13339', 'E133391', 'E133392', 'E133393', 'E133399', 'E13341', 'E133411', 'E133412', 'E133413', 'E133419', 'E13349', 'E133491', 'E133492', 'E133493', 'E133499', 'E13351', 'E133511', 'E133512', 'E133513', 'E133519', 'E133521', 'E133522', 'E133523', 'E133529', 'E133531', 'E133532', 'E133533', 'E133539', 'E133541', 'E133542', 'E133543', 'E133549', 'E133551', 'E133552', 'E133553', 'E133559', 'E13359', 'E133591', 'E133592', 'E133593', 'E133599', 'E1336', 'E1337X1', 'E1337X2', 'E1337X3', 'E1337X9', 'E1339', 'E1340', 'E1341', 'E1342', 'E1343', 'E1344', 'E1349', 'E1351', 'E1352', 'E1359', 'E13610', 'E13618', 'E13620', 'E13621', 'E13622', 'E13628', 'E13630', 'E13638', 'E13641', 'E13649', 'E1365', 'E1369', 'E138', 'E139', 'O24011', 'O24012', 'O24013', 'O24019', 'O2402', 'O2403', 'O24111', 'O24112', 'O24113', 'O24119', 'O2412', 'O2413', 'O24311', 'O24312', 'O24313', 'O24319', 'O2432', 'O2433', 'O24410', 'O24414', 'O24415', 'O24419', 'O2442', 'O2443', 'O24425', 'O24429', 'O24430', 'O24434', 'O24435', 'O24439', 'O24811', 'O24812', 'O24813', 'O24819', 'O2482', 'O2483', 'O24911', 'O24912', 'O24913', 'O24919', 'O2492', 'O2493']

Table S1, continued

Table 1, Continued

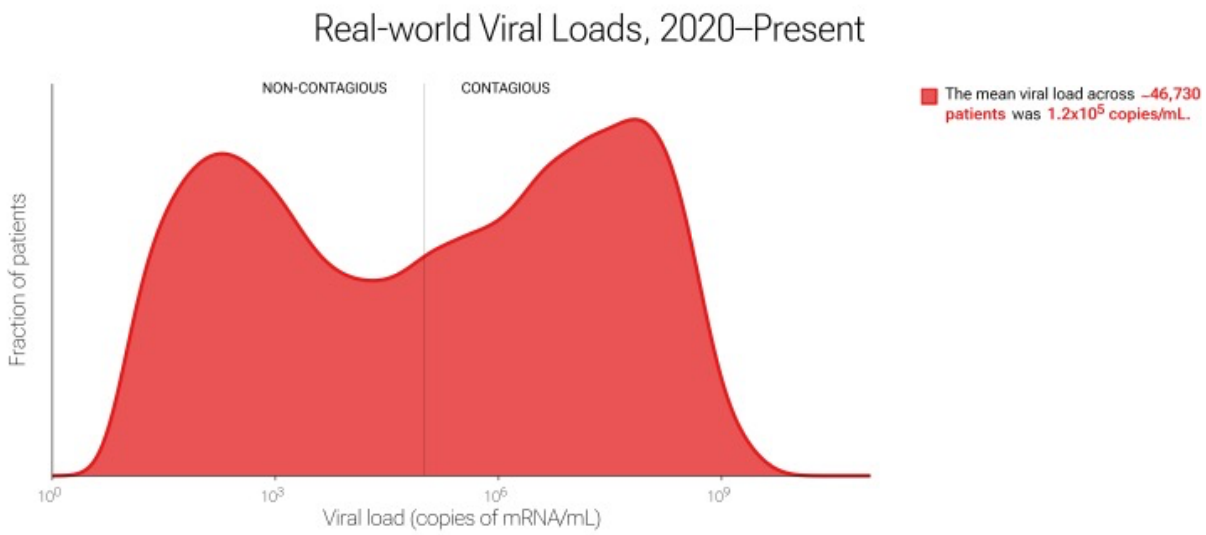
Checkbox name	Notes
Disabilities	ICD10 codes: ['G041', 'G800', 'G801', 'G802', 'G808', 'G809', 'G8100', 'G8101', 'G8102', 'G8103', 'G8104', 'G8110', 'G8111', 'G8112', 'G8113', 'G8114', 'G8190', 'G8191', 'G8192', 'G8193', 'G8194', 'G8220', 'G8221', 'G8222', 'G8250', 'G8251', 'G8252', 'G8253', 'G8254', 'G830', 'G8310', 'G8311', 'G8312', 'G8313', 'G8314', 'G8320', 'G8321', 'G8322', 'G8323', 'G8324', 'G8330', 'G8331', 'G8332', 'G8333', 'G8334', 'G834', 'G835', 'G8381', 'G8382', 'G8383', 'G8384', 'G8389', 'G839', 'I69031', 'I69032', 'I69033', 'I69034', 'I69039', 'I69041', 'I69042', 'I69043', 'I69044', 'I69049', 'I69051', 'I69052', 'I69053', 'I69054', 'I69059', 'I69061', 'I69062', 'I69063', 'I69064', 'I69065', 'I69069', 'I69131', 'I69132', 'I69133', 'I69134', 'I69139', 'I69141', 'I69142', 'I69143', 'I69144', 'I69149', 'I69151', 'I69152', 'I69153', 'I69154', 'I69159', 'I69161', 'I69162', 'I69163', 'I69164', 'I69165', 'I69169', 'I69231', 'I69232', 'I69233', 'I69234', 'I69239', 'I69241', 'I69242', 'I69243', 'I69244', 'I69249', 'I69251', 'I69252', 'I69253', 'I69254', 'I69259', 'I69261', 'I69262', 'I69263', 'I69264', 'I69265', 'I69269', 'I69331', 'I69332', 'I69333', 'I69334', 'I69339', 'I69341', 'I69342', 'I69343', 'I69344', 'I69349', 'I69351', 'I69352', 'I69353', 'I69354', 'I69359', 'I69361', 'I69362', 'I69363', 'I69364', 'I69365', 'I69369', 'I69831', 'I69832', 'I69833', 'I69834', 'I69839', 'I69841', 'I69842', 'I69843', 'I69844', 'I69849', 'I69851', 'I69852', 'I69853', 'I69854', 'I69859', 'I69861', 'I69862', 'I69863', 'I69864', 'I69865', 'I69869', 'I69931', 'I69932', 'I69933', 'I69934', 'I69939', 'I69941', 'I69942', 'I69943', 'I69944', 'I69949', 'I69951', 'I69952', 'I69953', 'I69954', 'I69959', 'I69961', 'I69962', 'I69963', 'I69964', 'I69965', 'I69969', 'Q', 'R532']
Heart conditions	ICD10 codes:['A1884', 'A3282', 'A3681', 'A381', 'A395', 'A5203', 'B2682', 'B332', 'B376', 'B5881', 'C452', 'D8685', 'G130', 'G712', 'G713', 'G720', 'G721', 'G722', 'G7249', 'G7281', 'G7289', 'G729', 'G737', 'I01', 'I02', 'I05', 'I06', 'I07', 'I08', 'I09', 'I0981', 'I11', 'I110', 'I13', 'I130', 'I20', 'I21', 'I22', 'I23', 'I24', 'I25', 'I252', 'I3', 'I4', 'I5', 'I501', 'I5020', 'I5021', 'I5022', 'I5023', 'I5030', 'I5031', 'I5032', 'I5033', 'I5040', 'I5041', 'I5042', 'I5043', 'I50810', 'I50811', 'I50812', 'I50813', 'I50814', 'I5082', 'I5083', 'I5084', 'I5089', 'I509', 'I5181', 'I70', 'I9713', 'I97130', 'I97131', 'J1082', 'J1182', 'O101', 'O29121', 'O29122', 'O29123', 'O29129', 'O903', 'OO2912', 'Q2', 'R570', 'S26', 'T82', 'Z95', 'Z95811', 'Z95812']
Immune status	Immunosuppressed vs. immunocompetent. Immunosuppresed if most recent T-cell subset analysis report, their absolute CD4 cell count was <100 cells/μl; they had a diagnosis of either lymphoma or leukemia associated with an encounter either before the PCR test or within 60 days of the PCR test; they were on any of the following drugs on an ongoing basis, prescribed prior to the PCR test and with enough refills to include the time up to 30 days prior to the PCR test: abatacept, adalimumab, anakinra, azathioprine, basiliximab, budesonide, certolizumab, cyclosporine, daclizumab, dexamethasone, everolimus, etanercept, golimumab, infliximab, ixekizumab, leflunomide, lenalidomide, methotrexate, mycophenolate, natalizumab, pomalidomide, prednisone, rituximab, secukinumab, serolimus, tacrolimus, tocilizumab, tofacitinib, ustekinumab, or vedolizumab. Otherwise, immunocompetent
Liver disease	ICD10 codes: ['A5145', 'A5274', 'B180', 'B181', 'B182', 'B188', 'B189', 'B190', 'B1910', 'B1911', 'B1920', 'B1921', 'B199', 'B251', 'B581', 'I8500', 'I8501', 'I8510', 'I8511', 'I864', 'K700', 'K702', 'K7030', 'K7031', 'K7040', 'K7041', 'K709', 'K713', 'K714', 'K7150', 'K7151', 'K716', 'K717', 'K718', 'K7210', 'K7211', 'K7290', 'K7291', 'K730', 'K731', 'K732', 'K738', 'K739', 'K740', 'K7400', 'K7401', 'K7402', 'K741', 'K742', 'K743', 'K744', 'K745', 'K7460', 'K7469', 'K751', 'K752', 'K753', 'K754', 'K7581', 'K7589', 'K759', 'K760', 'K761', 'K762', 'K763', 'K764', 'K765', 'K766', 'K767', 'K7681', 'K7689', 'K769', 'K77', 'K9182', 'Z944']
Mental health conditions	ICD10 codes:['F060', 'F061', 'F062', 'F0630', 'F0631', 'F0632', 'F0633', 'F0634', 'F11150', 'F11151', 'F11159', 'F11250', 'F11251', 'F11259', 'F11950', 'F11951', 'F11959', 'F12150', 'F12151', 'F12159', 'F12250', 'F12251', 'F12259', 'F12950', 'F12951', 'F12959', 'F13150', 'F13151', 'F13159', 'F13250', 'F13251', 'F13259', 'F13950', 'F13951', 'F13959', 'F14150', 'F14151', 'F14159', 'F14250', 'F14251', 'F14259', 'F14950', 'F14951', 'F14959', 'F15150', 'F15151', 'F15159', 'F15250', 'F15251', 'F15259', 'F15950', 'F15951', 'F15959', 'F16150', 'F16151', 'F16159', 'F16250', 'F16251', 'F16259', 'F16950', 'F16951', 'F16959', 'F18150', 'F18151', 'F18159', 'F18250', 'F18251', 'F18259', 'F18950', 'F18951', 'F18959', 'F19150', 'F19151', 'F19159', 'F19250', 'F19251', 'F19259', 'F19950', 'F19951', 'F19959', 'F200', 'F201', 'F202', 'F203', 'F205', 'F2081', 'F2089', 'F209', 'F21', 'F22', 'F23', 'F24', 'F250', 'F251', 'F258', 'F259', 'F28', 'F29', 'F3010', 'F3011', 'F3012', 'F3013', 'F302', 'F303', 'F304', 'F308', 'F309', 'F310', 'F3110', 'F3111', 'F3112', 'F3113', 'F312', 'F3130', 'F3131', 'F3132', 'F314', 'F315', 'F3160', 'F3161', 'F3162', 'F3163', 'F3164', 'F3170', 'F3171', 'F3172', 'F3173', 'F3174', 'F3175', 'F3176', 'F3177', 'F3178', 'F3181', 'F3189', 'F319', 'F320', 'F321', 'F322', 'F323', 'F324', 'F325', 'F328', 'F3281', 'F3289', 'F329', 'F32A', 'F330', 'F331', 'F332', 'F333', 'F3340', 'F3341', 'F3342', 'F338', 'F339', 'F340', 'F341', 'F348', 'F3481', 'F3489', 'F349', 'F39', 'F4489', 'F843']
Neighborhood income	median household income by zip code from the 2020 Census (2020 American Community Survey). URL structure (replace <ZIPCODE> with an actual zipcode): https://api.census.gov/data/2020/acs/acs5/subject?get=NAME,S1901_C01_012E&for=zip%20code%20tabulation%20area:<ZIPCODE> . In the text that this URL will return, S1901_C01_012E is the column name for median household income
Neurological disorders	ICD10 codes:['E7500', 'E7501', 'E7502', 'E7509', 'E7510', 'E7511', 'E7519', 'E7523', 'E7525', 'E7526', 'E7529', 'E754', 'F0150', 'F0151', 'F0280', 'F0281', 'F0390', 'F0391', 'F05', 'F842', 'G08', 'G10', 'G110', 'G111', 'G1110', 'G1111', 'G1119', 'G112', 'G113', 'G114', 'G118', 'G119', 'G120', 'G121', 'G1220', 'G1221', 'G1222', 'G1223', 'G1224', 'G1225', 'G1229', 'G128', 'G129', 'G131', 'G132', 'G138', 'G20', 'G210', 'G2111', 'G2119', 'G212', 'G213', 'G214', 'G218', 'G219', 'G230', 'G231', 'G232', 'G238', 'G239', 'G2409', 'G241', 'G242', 'G248', 'G254', 'G255', 'G2570', 'G2571', 'G2579', 'G2581', 'G2582', 'G2583', 'G2589', 'G259', 'G26', 'G300', 'G301', 'G308', 'G309', 'G3101', 'G3109', 'G311', 'G312', 'G3181', 'G3182', 'G3183', 'G3185', 'G3189', 'G319', 'G320', 'G3281', 'G3289', 'G35', 'G360', 'G368', 'G369', 'G370', 'G371', 'G372', 'G373', 'G374', 'G375', 'G378', 'G379', 'G40001', 'G40009', 'G40011', 'G40019', 'G40101', 'G40109', 'G40111', 'G40119', 'G40201', 'G40209', 'G40211', 'G40219', 'G40301', 'G40309', 'G40311', 'G40319', 'G40401', 'G40409', 'G40411', 'G40419', 'G4042', 'G40501', 'G40509', 'G40801', 'G40802', 'G40803', 'G40804', 'G40811', 'G40812', 'G40813', 'G40814', 'G40821', 'G40822', 'G40823', 'G40824', 'G40833', 'G40834', 'G4089', 'G40901', 'G40909', 'G40911', 'G40919', 'G40A01', 'G40A09', 'G40A11', 'G40A19', 'G40B01', 'G40B09', 'G40B11', 'G40B19', 'G47411', 'G47419', 'G47421', 'G47429', 'G803', 'G890', 'G910', 'G911', 'G912', 'G913', 'G914', 'G918', 'G919', 'G930', 'G9340', 'G9341', 'G9349', 'G935', 'G936', 'G937', 'G9381', 'G9382', 'G9389', 'G939', 'G94', 'O99350', 'O99351', 'O99352', 'O99353', 'O99354', 'O99355', 'P9160', 'P9161', 'P9162', 'P9163', 'R561', 'R569']
Outcome	Died from COVID-19 (causal), Died with COVID-19 (incidental), Survived
Patient location	Inpatient, Outpatient, Emergency Room, or Institutional (sent from another hospital)
Peptic ulcer disease	ICD10 codes: ['K250', 'K251', 'K252', 'K253', 'K254', 'K255', 'K256', 'K257', 'K259', 'K260', 'K261', 'K262', 'K263', 'K264', 'K265', 'K266', 'K267', 'K269', 'K270', 'K271', 'K272', 'K273', 'K274', 'K275', 'K276', 'K277', 'K279', 'K280', 'K281', 'K282', 'K283', 'K284', 'K285', 'K286', 'K287', 'K289']
Peripheral vascular disease	ICD10 codes: ['A5200', 'A5201', 'A5202', 'A5209', 'I700', 'I701', 'I70201', 'I70202', 'I70203', 'I70208', 'I70209', 'I70211', 'I70212', 'I70213', 'I70218', 'I70219', 'I70221', 'I70222', 'I70223', 'I70228', 'I70229', 'I70231', 'I70232', 'I70233', 'I70234', 'I70235', 'I70238', 'I70239', 'I70241', 'I70242', 'I70243', 'I70244', 'I70245', 'I70248', 'I70249', 'I7025', 'I70261', 'I70262', 'I70263', 'I70268', 'I70269', 'I70291', 'I70292', 'I70293', 'I70298', 'I70299', 'I70301', 'I70302', 'I70303', 'I70308', 'I70309', 'I70311', 'I70312', 'I70313', 'I70318', 'I70319', 'I70321', 'I70322', 'I70323', 'I70328', 'I70329', 'I70331', 'I70332', 'I70333', 'I70334', 'I70335', 'I70338', 'I70339', 'I70341', 'I70342', 'I70343', 'I70344', 'I70345', 'I70348', 'I70349', 'I7035', 'I70361', 'I70362', 'I70363', 'I70368', 'I70369', 'I70391', 'I70392', 'I70393', 'I70398', 'I70399', 'I70401', 'I70402', 'I70403', 'I70408', 'I70409', 'I70411', 'I70412', 'I70413', 'I70418', 'I70419', 'I70421', 'I70422', 'I70423', 'I70428', 'I70429', 'I70431', 'I70432', 'I70433', 'I70434', 'I70435', 'I70438', 'I70439', 'I70441', 'I70442', 'I70443', 'I70444', 'I70445', 'I70448', 'I70449', 'I7045', 'I70461', 'I70462', 'I70463', 'I70468', 'I70469', 'I70491', 'I70492', 'I70493', 'I70498', 'I70499', 'I70501', 'I70502', 'I70503', 'I70508', 'I70509', 'I70511', 'I70512', 'I70513', 'I70518', 'I70519', 'I70521', 'I70522', 'I70523', 'I70528', 'I70529', 'I70531', 'I70532', 'I70533', 'I70534', 'I70535', 'I70538', 'I70539', 'I70541', 'I70542', 'I70543', 'I70544', 'I70545', 'I70548', 'I70549', 'I7055', 'I70561', 'I70562', 'I70563', 'I70568', 'I70569', 'I70591', 'I70592', 'I70593', 'I70598', 'I70599', 'I70601', 'I70602', 'I70603', 'I70608', 'I70609', 'I70611', 'I70612', 'I70613', 'I70618', 'I70619', 'I70621', 'I70622', 'I70623', 'I70628', 'I70629', 'I70631', 'I70632', 'I70633', 'I70634', 'I70635', 'I70638', 'I70639', 'I70641', 'I70642', 'I70643', 'I70644', 'I70645', 'I70648', 'I70649', 'I7065', 'I70661', 'I70662', 'I70663', 'I70668', 'I70669', 'I70691', 'I70692', 'I70693', 'I70698', 'I70699', 'I70701', 'I70702', 'I70703', 'I70708', 'I70709', 'I70711', 'I70712', 'I70713', 'I70718', 'I70719', 'I70721', 'I70722', 'I70723', 'I70728', 'I70729', 'I70731', 'I70732', 'I70733', 'I70734', 'I70735', 'I70738', 'I70739', 'I70741', 'I70742', 'I70743', 'I70744', 'I70745', 'I70748', 'I70749', 'I7075', 'I70761', 'I70762', 'I70763', 'I70768', 'I70769', 'I70791', 'I70792', 'I70793', 'I70798', 'I70799', 'I708', 'I7090', 'I7091', 'I7092', 'I7100', 'I7101', 'I7102', 'I7103', 'I711', 'I712', 'I713', 'I714', 'I715', 'I716', 'I718', 'I719', 'I720', 'I721', 'I722', 'I723', 'I724', 'I725', 'I726', 'I728', 'I729', 'I7301', 'I731', 'I7381', 'I7389', 'I739', 'I7401', 'I7409', 'I7410', 'I7411', 'I7419', 'I742', 'I743', 'I744', 'I745', 'I748', 'I749', 'I75011', 'I75012', 'I75013', 'I75019', 'I75021', 'I75022', 'I75023', 'I75029', 'I7581', 'I7589', 'I770', 'I771', 'I772', 'I773', 'I774', 'I775', 'I776', 'I7770', 'I7771', 'I7772', 'I7773', 'I7774', 'I7775', 'I7776', 'I7777', 'I7779', 'I77810', 'I77811', 'I77812', 'I77819', 'I7789', 'I779', 'I780', 'I781', 'I788', 'I789', 'I790', 'I791', 'I798', 'K551', 'Z95820', 'Z95828']
Pregnancy status	Pregnant or Recently delivered, Not pregnant. (males, individuals under 13 or over 56, and recent test results excluded from results)
Presumed variant	Early variants, delta, omicron
Pulmonary disease	ICD10 codes: ['J410', 'J411', 'J418', 'J42', 'J430', 'J431', 'J432', 'J438', 'J439', 'J440', 'J441', 'J449', 'J470', 'J471', 'J479', 'J60', 'J61', 'J620', 'J628', 'J630', 'J631', 'J632', 'J633', 'J634', 'J635', 'J636', 'J64', 'J65', 'J660', 'J661', 'J662', 'J668', 'J670', 'J671', 'J672', 'J673', 'J674', 'J675', 'J676', 'J677', 'J678', 'J679', 'J684', 'J701', 'J703']
Race/Ethnicity	self-reported. Allowed values: Black, White, Hispanic, Asian + Pacific Islander, Other + Unknown.
Remdesivir	matches to fuzzy-match case-insensitive searching for strings beginning "remd"
Renal disease	ICD10 codes: ['I120', 'I1311', 'I132', 'N183', 'N1830', 'N1831', 'N1832', 'N184', 'N185', 'N186', 'N189', 'N19', 'Z4901', 'Z4902', 'Z4931', 'Z4932', 'Z9115', 'Z940', 'Z992']
Sex	Female vs. male
Sickle cell & thalassemia	ICD10 codes: ["D56","D57"]
Smoking status	Current, former, never
Substance abuse	ICD10 codes: ['F1010', 'F1011', 'F10120', 'F10121', 'F10129', 'F10130', 'F10131', 'F10132', 'F10139', 'F1014', 'F10150', 'F10151', 'F10159', 'F10180', 'F10181', 'F10182', 'F10188', 'F1019', 'F1020', 'F1021', 'F10220', 'F10221', 'F10229', 'F10230', 'F10231', 'F10232', 'F10239', 'F1024', 'F10250', 'F10251', 'F10259', 'F1026', 'F1027', 'F10280', 'F10281', 'F10282', 'F10288', 'F1029', 'F1094', 'F10950', 'F10951', 'F10959', 'F1096', 'F1097', 'F10980', 'F11', 'F1110', 'F1111', 'F11120', 'F11121', 'F11122', 'F11129', 'F1113', 'F1114', 'F11181', 'F11182', 'F11188', 'F1119', 'F1120', 'F1121', 'F11220', 'F11221', 'F11222', 'F11229', 'F1123', 'F1124', 'F11281', 'F11282', 'F11288', 'F1129', 'F12', 'F1210', 'F1211', 'F12120', 'F12121', 'F12122', 'F12129', 'F1213', 'F12180', 'F12188', 'F1219', 'F1220', 'F1221', 'F12220', 'F12221', 'F12222', 'F12229', 'F1223', 'F12280', 'F12288', 'F1229', 'F1310', 'F1311', 'F13120', 'F13121', 'F13129', 'F13130', 'F13131', 'F13132', 'F13139', 'F1314', 'F13180', 'F13181', 'F13182', 'F13188', 'F1319', 'F1320', 'F1321', 'F13220', 'F13221', 'F13229', 'F13230', 'F13231', 'F13232', 'F13239', 'F1324', 'F1326', 'F1327', 'F13280', 'F13281', 'F13282', 'F13288', 'F1329', 'F14', 'F1410', 'F1411', 'F14120', 'F14121', 'F14122', 'F14129', 'F1413', 'F1414', 'F14180', 'F14181', 'F14182', 'F14188', 'F1419', 'F1420', 'F1421', 'F14220', 'F14221', 'F14222', 'F14229', 'F1423', 'F1424', 'F14280', 'F14281', 'F14282', 'F14288', 'F1429', 'F1510', 'F1511', 'F15120', 'F15121', 'F15122', 'F15129', 'F1513', 'F1514', 'F15180', 'F15181', 'F15182', 'F15188', 'F1519', 'F1520', 'F1521', 'F15220', 'F15221', 'F15222', 'F15229', 'F1523', 'F1524', 'F15280', 'F15281', 'F15282', 'F15288', 'F1529', 'F16', 'F1610', 'F1611', 'F16120', 'F16121', 'F16122', 'F16129', 'F1614', 'F16180', 'F16183', 'F16188', 'F1619', 'F1620', 'F1621', 'F16220', 'F16221', 'F16229', 'F1624', 'F16280', 'F16283', 'F16288', 'F1629', 'F1810', 'F1811', 'F18120', 'F18121', 'F18129', 'F1814', 'F1817', 'F18180', 'F18188', 'F1819', 'F1820', 'F1821', 'F18220', 'F18221', 'F18229', 'F1824', 'F1827', 'F18280', 'F18288', 'F1829', 'F19', 'F1910', 'F1911', 'F19120', 'F19121', 'F19122', 'F19129', 'F19130', 'F19131', 'F19132', 'F19139', 'F1914', 'F1916', 'F1917', 'F19180', 'F19181', 'F19182', 'F19188', 'F1919', 'F1920', 'F1921', 'F19220', 'F19221', 'F19222', 'F19229', 'F19230', 'F19231', 'F19232', 'F19239', 'F1924', 'F1926', 'F1927', 'F19280', 'F19281', 'F19282', 'F19288', 'F1929', 'G621', 'I426', 'K2920', 'K2921', 'K7010', 'K7011', 'O99310', 'O99311', 'O99312', 'O99313', 'O99314', 'O99315', 'O99320', 'O99321', 'O99322', 'O99323', 'O99324', 'O99325']
Tocilizumab	matches to case-insensitive searches for the string "tocilizumab"
Transplanted organ and tissue status	ICD10 codes: ["Z94"]
Ventilation assistance	CPT4 codes: ["94002", "94003", "94640", "94645"]

562 Figure 1: Patient characteristics available for defining patient groups and subgroups



563

564 Figure 2: Overall bimodal distribution of viral loads

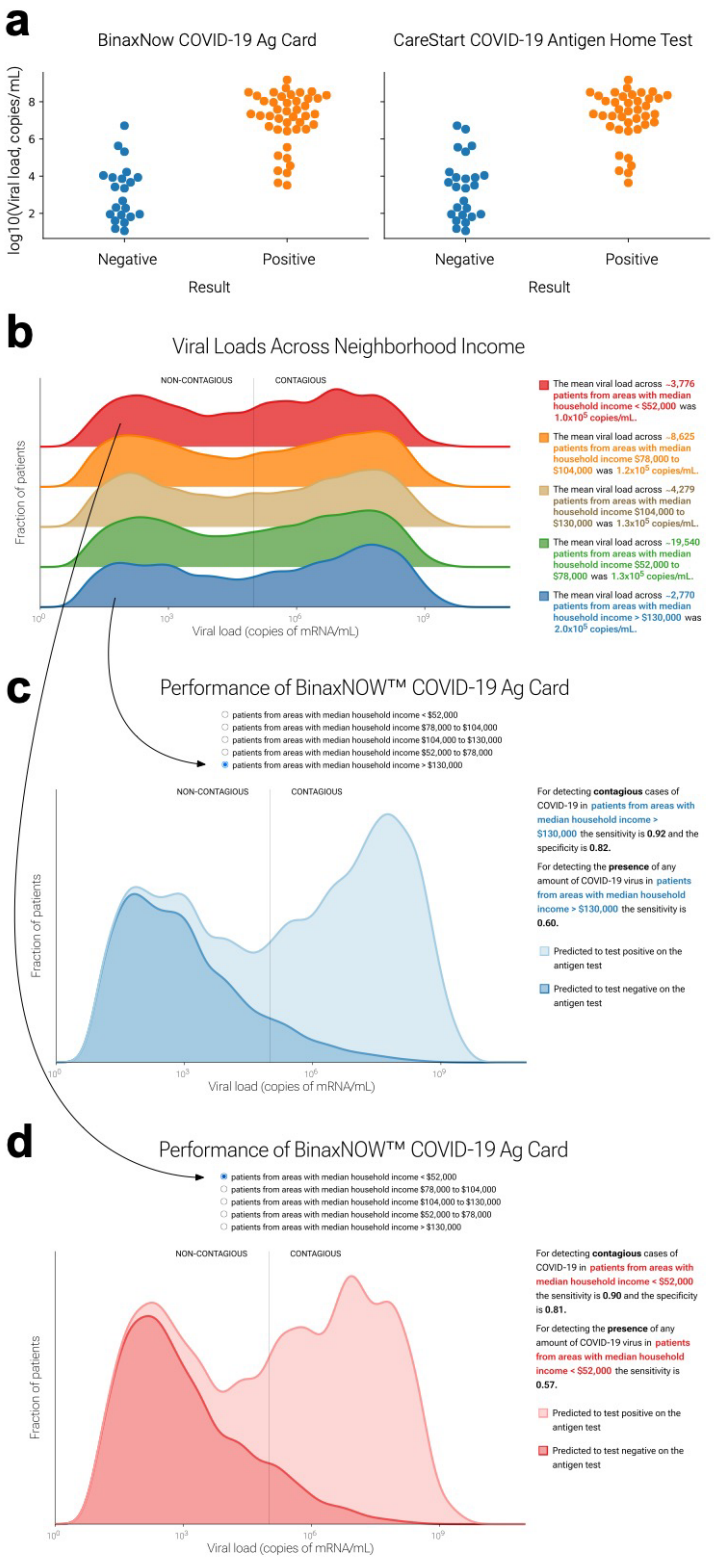


565

566 Figure 3: Viral load comparisons by remdesivir treatment, patient presentation, and outcome

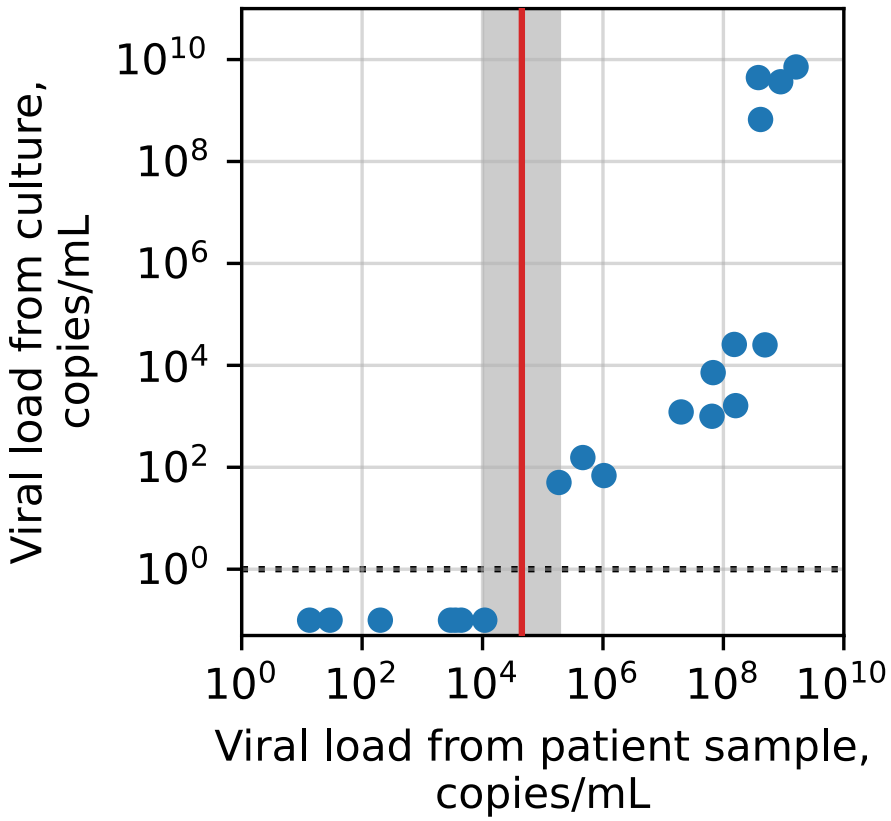
567

568 Figure 4: Antigen test results from head-to-head trial and performance on patient subgroups



569

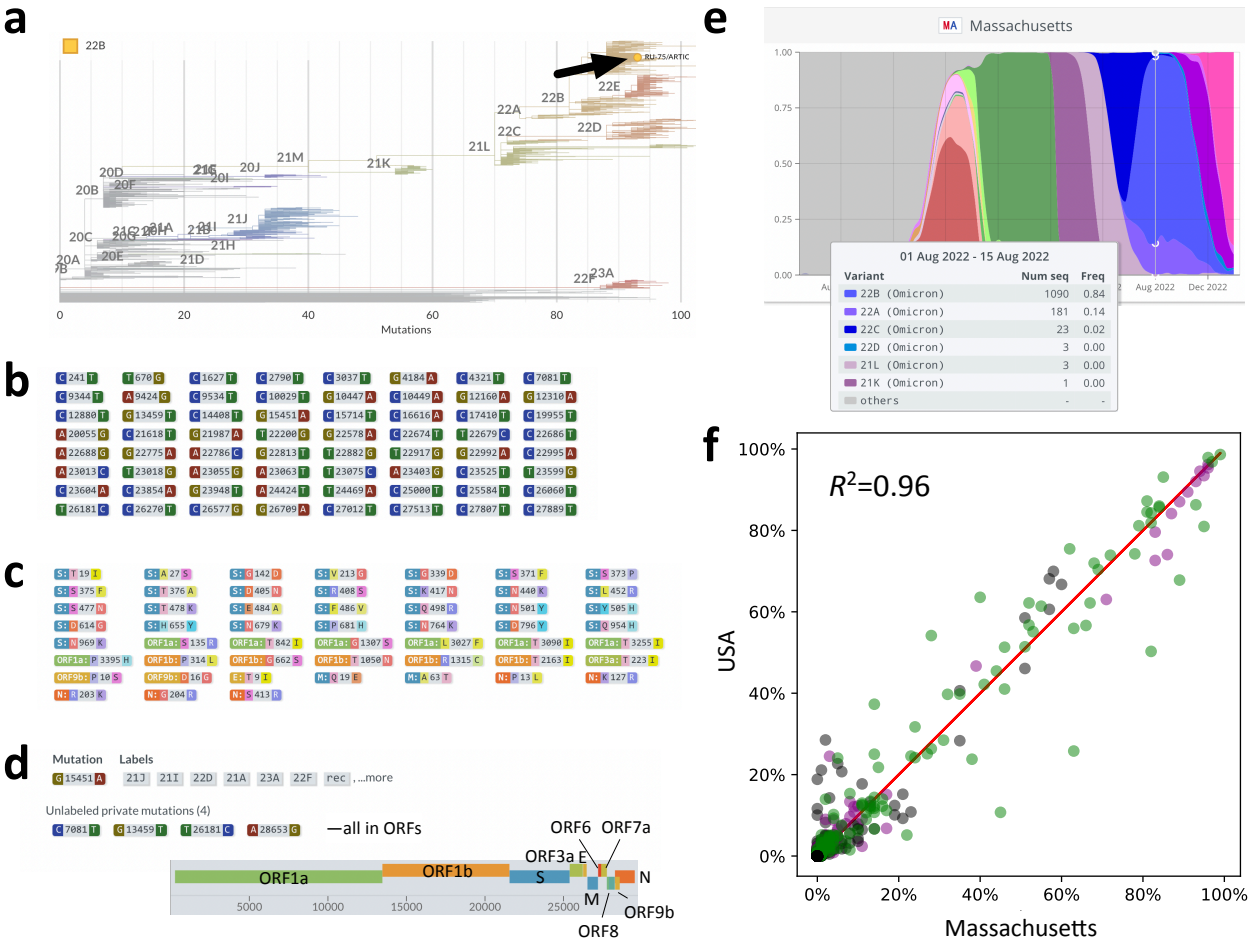
570 Figure 5: Determination of the contagiousness threshold for omicron-era SARS-CoV-2 strains



571

572 Figure 6: Sequencing late-2022 strain and generalizability of Massachusetts-level results to the

573 United States as a whole



574

References

1. National Research Council, Division on Earth and Life Studies, Board on Life Sciences, Committee on a Framework for Developing a New Taxonomy of Disease (2011). Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease (National Academies Press).

2. Arnaout, R., Lee, R.A., Lee, G.R., Callahan, C., Cheng, A., Yen, C.F., Smith, K.P., Arora, R., and Kirby, J.E. (2021). The Limit of Detection Matters: The Case for Benchmarking Severe Acute Respiratory Syndrome Coronavirus 2 Testing. Clin Infect Dis. 10.1093/cid/ciaa1382.

3. Wild, D. ed. (2013). The immunoassay handbook: theory and applications of ligand binding, ELISA, and related techniques 4th ed. (Elsevier).

4. Hill, E.D., Yilmaz, F., Callahan, C., Cheng, A., Braun, J., and Arnaout, R. (2022). *ct2vl* : Converting Ct Values to Viral Loads for SARS-CoV-2 RT-qPCR Test Results (Microbiology) 10.1101/2022.06.20.496929.

5. Callahan Cody, Lee Rose A., Lee Ghee Rye, Zulauf Kate, Kirby James E., Arnaout Ramy, and Miller Melissa B. Nasal Swab Performance by Collection Timing, Procedure, and Method of Transport for Patients with SARS-CoV-2. Journal of Clinical Microbiology *59*, e00569-21. 10.1128/JCM.00569-21.

6. Callahan, C., Ditelberg, S., Dutta, S., Littlehale, N., Cheng, A., Kupczewski, K., McVay, D., Riedel, S., Kirby, J.E., and Arnaout, R. (2021). Saliva is Comparable to Nasopharyngeal Swabs for Molecular Detection of SARS-CoV-2. Microbiol Spectr *9*, e0016221. 10.1128/Spectrum.00162-21.

7. Pilarowski, G., Lebel, P., Sunshine, S., Liu, J., Crawford, E., Marquez, C., Rubio, L., Chamie, G., Martinez, J., Peng, J., et al. (2021). Performance Characteristics of a Rapid Severe Acute Respiratory Syndrome Coronavirus 2 Antigen Detection Assay at a Public Plaza Testing Site in San Francisco. J Infect Dis *223*, 1139–1144. 10.1093/infdis/jiaa802.

8. Mina, M.J., Parker, R., and Larremore, D.B. (2020). Rethinking Covid-19 Test Sensitivity - A Strategy for Containment. N Engl J Med *383*, e120. 10.1056/NEJMp2025631.

9. Stanley, S., Hamel, D.J., Wolf, I.D., Riedel, S., Dutta, S., Contreras, E., Callahan, C.J., Cheng, A., Arnaout, R., Kirby, J.E., et al. (2022). Limit of Detection for Rapid Antigen Testing of the SARS-CoV-2 Omicron and Delta Variants of Concern Using Live-Virus Culture. J Clin Microbiol *60*, e00140-22. 10.1128/jcm.00140-22.

10. Kirby, J.E., Riedel, S., Dutta, S., Arnaout, R., Cheng, A., Ditelberg, S., Hamel, D.J., Chang, C.A., and Kanki, P.J. (2022). SARS-CoV-2 Antigen Tests Predict Infectivity based on viral culture: comparison of antigen, PCR viral load, and viral culture testing on a large Sample Cohort. Clinical Microbiology and Infection, S1198743X22003743. 10.1016/j.cmi.2022.07.010.

11. Charlson, M.E., Pompei, P., Ales, K.L., and MacKenzie, C.R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis *40*, 373–383. 10.1016/0021-9681(87)90171-8.

12. Greenberg, J.A., Hohmann, S.F., Hall, J.B., Kress, J.P., and David, M.Z. (2016). Validation of a Method to Identify Immunocompromised Patients with Severe Sepsis in Administrative Databases. Annals ATS *13*, 253–258. 10.1513/AnnalsATS.201507-415BC.

13. Knight, S.R., Ho, A., Pius, R., Buchan, I., Carson, G., Drake, T.M., Dunning, J., Fairfield, C.J., Gamble, C., Green, C.A., et al. (2020). Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ *370*, m3339. 10.1136/bmj.m3339.

14. Elixhauser, A., Steiner, C., Harris, D.R., and Coffey, R.M. (1998). Comorbidity measures for use with administrative data. Med Care *36*, 8–27. 10.1097/00005650-199801000-00004.

15. Shain, E.B., and Clemens, J.M. (2008). A new method for robust quantitative and qualitative analysis of real-time PCR. Nucleic Acids Res *36*, e91. 10.1093/nar/gkn408.

16. Kirby, J.E., Cheng, A., Cleveland, M.H., Degli-Angeli, E., DeMarco, C.T., Faron, M., Gallagher, T., Garlick, R.K., Goecker, E., Coombs, R.W., et al. (2022). A Multi-Institutional Study Benchmarking Cycle Threshold Values for Major Clinical SARS-CoV-2 RT-PCR Assays. 2022.06.22.22276072. 10.1101/2022.06.22.22276072.

17. CoVariants https://covariants.org/.

18. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research *12*, 2825–2830.

19. Wölfel, R., Corman, V.M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M.A., Niemeyer, D., Jones, T.C., Vollmar, P., Rothe, C., et al. (2020). Virological assessment of hospitalized patients with COVID-2019. Nature *581*, 465–469. 10.1038/s41586-020-2196-x.

20. La Scola, B., Le Bideau, M., Andreani, J., Hoang, V.T., Grimaldier, C., Colson, P., Gautret, P., and Raoult, D. (2020). Viral RNA load as determined by cell culture as a management tool for discharge of SARS-CoV-2 patients from infectious disease wards. Eur J Clin Microbiol Infect Dis *39*, 1059–1061. 10.1007/s10096-020-03913-9.

21. Huang, C.-G., Lee, K.-M., Hsiao, M.-J., Yang, S.-L., Huang, P.-N., Gong, Y.-N., Hsieh, T.-H., Huang, P.-W., Lin, Y.-J., Liu, Y.-C., et al. (2020). Culture-Based Virus Isolation To Evaluate Potential Infectivity of Clinical Specimens Tested for COVID-19. J Clin Microbiol *58*, e01068-20. 10.1128/JCM.01068-20.

22. Artic Network https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html.

23. Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A., et al. (2017). Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc *12*, 1261–1276. 10.1038/nprot.2017.066.

24. Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R.A. (2018). Nextstrain: real-time tracking of pathogen evolution. Bioinformatics *34*, 4121–4123. 10.1093/bioinformatics/bty407.

25. Pratt, J.W., and Gibbons, J.D. Concepts of nonparametric theory (Springer International Publishing).

26. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

27. Tukey, J. (1953). Multiple Comparisons. J. Am. Stat. Assoc. *48*, 624–625.

28. Morgan, A. COVIRAL webapp back-end. https://github.com/chhotii-alex/antigen-flask.

29. Morgan, A. COVIRAL webapp front-end. https://github.com/chhotii-alex/antigen-sensitivity.

30. Lamb, Y.N. (2020). Remdesivir: First Approval. Drugs *80*, 1355–1363. 10.1007/s40265-020-01378-w.

31. Ahmed, A., Song, Y., and Wadhera, R.K. (2022). Racial/Ethnic Disparities in Delaying or Not Receiving Medical Care During the COVID-19 Pandemic. J GEN INTERN MED *37*, 1341–1343. 10.1007/s11606-022-07406-7.

32. Watson, C. (2022). Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever. Nature Medicine *28*, 2–5. 10.1038/s41591-021-01654-6.

33. Rando, H.M., Boca, S.M., McGowan, L.D., Himmelstein, D.S., Robson, M.P., Rubinetti, V., Velazquez, R., Greene, C.S., and Gitter, A. (2021). An Open-Publishing Response to the COVID-19 Infodemic. CEUR Workshop Proc *2976*, 29–38.

34. Ross, J.S. (2021). Covid-19, open science, and the CVD-COVID-UK initiative. BMJ *373*, n898. 10.1136/bmj.n898.

671 35. Callahan, C.J., Lee, R., Zulauf, K.E., Tamburello, L., Smith, K.P., Previtera, J., Cheng, A., Green,
672 A., Azim, A.A., Yano, A., et al. (2020). Open Development and Clinical Validation Of Multiple
673 3D-Printed Nasopharyngeal Collection Swabs: Rapid Resolution of a Critical COVID-19
674 Testing Bottleneck. J Clin Microbiol, JCM.00876-20, jcm;JCM.00876-20v1.
675 10.1128/JCM.00876-20.

676 36. Arnaout, R.A. (2021). Cooperation under Pressure: Lessons from the COVID-19 Swab Crisis.
677 Journal of Clinical Microbiology *59*, e01239-21. 10.1128/JCM.01239-21.

678 37. Tse, E.G., Klug, D.M., and Todd, M.H. (2020). Open science approaches to COVID-19.
679 F1000Res *9*, 1043. 10.12688/f1000research.26084.1.

680 38. Perillat, L., and Baigrie, B.S. (2021). COVID-19 and the generation of novel scientific
681 knowledge: Evidence-based decisions and data sharing. J Eval Clin Pract *27*, 708–715.
682 10.1111/jep.13548.

683 39. Li, R., von Isenburg, M., Levenstein, M., Neumann, S., Wood, J., and Sim, I. (2021). COVID-19
684 trials: declarations of data sharing intentions at trial registration and at publication. Trials
685 *22*, 153. 10.1186/s13063-021-05104-z.

686 40. Gardener, A.D., Hick, E.J., Jacklin, C., Tan, G., Cashin, A.G., Lee, H., Nunan, D., Toomey, E.C.,
687 and Richards, G.C. (2022). Open science and conflict of interest policies of medical and
688 health sciences journals before and during the COVID-19 pandemic: A repeat cross-
689 sectional study: Open science policies of medical journals. JRSM Open *13*,
690 20542704221132139. 10.1177/20542704221132139.

691 41. Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E.,
692 Deforet, M., and Leyrat, C. (2021). Open science saves lives: lessons from the COVID-19
693 pandemic. BMC Med Res Methodol *21*, 117. 10.1186/s12874-021-01304-y.

694 42. Wood, A., Denholm, R., Hollings, S., Cooper, J., Ip, S., Walker, V., Denaxas, S., Akbari, A.,
695 Banerjee, A., Whiteley, W., et al. (2021). Linked electronic health records for research on a

nationwide cohort of more than 54 million people in England: data resource. BMJ *373*, n826. 10.1136/bmj.n826.

43. Rao, S.N., Manissero, D., Steele, V.R., and Pareja, J. (2020). A Systematic Review of the Clinical Utility of Cycle Threshold Values in the Context of COVID-19. Infect Dis Ther *9*, 573–586. 10.1007/s40121-020-00324-3.

44. Satlin, M.J., Chen, L., Gomez-Simmonds, A., Marino, J., Weston, G., Bhowmick, T., Seo, S.K., Sperber, S.J., Kim, A.C., Eilertson, B., et al. (2022). Impact of a Rapid Molecular Test for Klebsiella pneumoniae Carbapenemase and Ceftazidime-Avibactam Use on Outcomes After Bacteremia Caused by Carbapenem-Resistant Enterobacterales. Clinical Infectious Diseases *75*, 2066–2075. 10.1093/cid/ciac354.

45. Savela, E.S., Viloria Winnett, A., Romano, A.E., Porter, M.K., Shelby, N., Akana, R., Ji, J., Cooper, M.M., Schlenker, N.W., Reyes, J.A., et al. (2022). Quantitative SARS-CoV-2 Viral-Load Curves in Paired Saliva Samples and Nasal Swabs Inform Appropriate Respiratory Sampling Site and Analytical Test Sensitivity Required for Earliest Viral Detection. J Clin Microbiol *60*, e0178521. 10.1128/JCM.01785-21.

46. Viloria Winnett, A., Porter, M.K., Romano, A.E., Savela, E.S., Akana, R., Shelby, N., Reyes, J.A., Schlenker, N.W., Cooper, M.M., Carter, A.M., et al. (2022). Morning SARS-CoV-2 Testing Yields Better Detection of Infection Due to Higher Viral Loads in Saliva and Nasal Swabs upon Waking. Microbiol Spectr *10*, e0387322. 10.1128/spectrum.03873-22.

47. Brümmer, L.E., Katzenschlager, S., Gaeddert, M., Erdmann, C., Schmitz, S., Bota, M., Grilli, M., Larmann, J., Weigand, M.A., Pollock, N.R., et al. (2021). Accuracy of novel antigen rapid diagnostics for SARS-CoV-2: A living systematic review and meta-analysis. PLoS Med *18*, e1003735. 10.1371/journal.pmed.1003735.

48. Lee, R.A., Herigon, J.C., Benedetti, A., Pollock, N.R., and Denkinger, C.M. (2021). Performance of Saliva, Oropharyngeal Swabs, and Nasal Swabs for SARS-CoV-2 Molecular

721 Detection: A Systematic Review and Meta-analysis. Journal of Clinical Microbiology.
722 10.1128/JCM.02881-20.

723 49. El Emam, K. (2013). Guide to the de-Identification of Personal Health Information (Auerbach
724 Publishers, Incorporated).

725 50. El Emam, K., Rodgers, S., and Malin, B. (2015). Anonymising and sharing individual patient
726 data. BMJ *350*, h1139. 10.1136/bmj.h1139.

727