**Supplementary Information**

**Supplementary Methods**

*Setting*

United States (U.S.) Department of Veterans Affairs (VA), Veterans Health Administration (VHA) healthcare facilities.

*Description of the National Healthcare Safety Network (NHSN) COVID-19 acute care module data element reviewed in this project*

This analysis focused on VHA facility data reported to the Centers for Disease Control and Prevention (CDC) NHSN COVID-19 Acute Care Module for the data element pertaining to the daily count of patients hospitalized with COVID-19 (hereafter referred to as the “hospitalized” count). The “hospitalized” data element description and instructions are listed below as they existed at the time of this review conducted in June 2020.

NHSN Data Element Name: Hospitalized

NHSN Data Element Description: Patients currently hospitalized in an inpatient bed who have suspecteda or confirmedb COVID-19.

NHSN Instructions: Enter the number of patients hospitalized in an inpatient bed at the time the data is collected who have suspected or confirmed COVID-19.

Footnotes:

a Suspected: A patient without COVID-19 (SARS CoV-2) laboratory viral test indicating current infection (note, this does not include serology testing for antibody) who in accordance with CDC’s Evaluating and Testing Persons for Coronavirus Disease 2019 (COVID-19), has signs and symptoms compatible with COVID-19. Most patients with confirmed COVID-19 have fever and/or symptoms of acute respiratory illness (cough, shortness of breath, difficulty breathing) but some people may present with other symptoms such as chills, repeated shaking with chills, muscle pain, new loss of taste or smell, headache or sore throat

b Confirmed: A patient with a positive COVID-19 (SARS CoV-2) laboratory viral test indicating current infection (Note this does not include serology testing for antibody.)

*Recruitment of facilities for comparison of computer-extracted and facility-reported “hospitalized” patient counts*

VHA Central Office leadership gave approval for this validation project, including approaching facility leadership at all 170 VHA medical facilities in the U.S. and territories to volunteer to participate. Volunteer facilities reported data to the authors corresponding to a two-week period in June 2020 (weekdays only, for a total of 10 days). These facility-reported data were compared to the computer-extracted data reported by VHA to NHSN for the same 10-day period.

*Data Source: Computer Extraction Tool*

The Department of Veterans Affairs (VA) National Surveillance Tool, initiated in March 2020 for centralized COVID-19 data collection, collates data from the bed management system (shows patient location and whether the patient is suspected or confirmed to have COVID-19), VHA health factors (small messages representing certain occurrences) that reflect the results of COVID-19 screening questions, and laboratory results. For NHSN COVID-19 Acute Care Module reporting, data were pulled once daily from resources managed in the VA Corporate Data Warehouse and processed through algorithms that approximated the CDC questionnaire items. For example, the presence of a health factor for a positive COVID-19 screening survey, a pending test, or a bed management flag indicating a Patient Under Investigation were counted as a suspected case. Those with health factors indicating a positive test, a positive VHA laboratory test, or a bed management flag indicating confirmed status were counted as a confirmed case.  These data were then aggregated per facility daily into a summary table to generate output for NHSN submittal.

*Data Source: Facility Data Collection*

Facilities that chose to participate in this project appointed a Point of Contact (POC) for data collection. A Microsoft Excel™ template was provided to each POC for reporting of daily hospitalized patient counts and patient personally identifiable information (PII) for the 10-day reporting period. The NHSN definition for hospitalized patients was provided in the Excel™ workbook and POCs were instructed to use whatever method would be used at the facility to determine the hospitalized patient count for each day in accordance with the NHSN definition. The POCs were instructed to report whether each patient counted on each day was considered to be a confirmed or suspected COVID-19 case. Prior to initiation of data collection, POCs were required to attend an hour-long virtual training session for project organizers to explain the data collection sheets and how to determine the daily patient counts. The slides used for the training session and a Frequently Asked Questions information sheet were provided to the POCs.

*Comparison of Data between Computer-extracted and Facility-reported Data Sources*

For the 10 weekdays of the comparison project, the computer extraction compiled the number of cases that met the definition for suspected or confirmed hospitalized patients at each facility at 10:00 AM local time. Likewise, POCs at the participating facilities were instructed to count patients with suspected and confirmed COVID-19 hospitalized as of 10:00 AM local time. In both systems, patients were counted on each day they were hospitalized. Results of the hospitalized patient counts from the Facility Reports and the Computer Extracts for all 36 participating facilities were compiled into a master Excel™ spreadsheet (by authors L.F. and S.G.) for determination of hospitalized patient count matching for each facility-day.

For analyses on a facility level, another Excel™ template was created for each facility for comparison of individual patient PII from Facility Reports and Computer Extracts to determine how well the patients contributing to the daily counts matched. Completion of these templates was done by authors S.G., M.E., B.M, K.L and L.F.

For the facilities in which there were zero days or one day of matching counts between the data sources, chart reviews were done (by authors SG, M.E, and B.M), using the facility Excel™ templates, to determine if the patients should have been counted based on the NHSN reporting definitions. Information collected from chart reviews included: whether or not a COVID-19 test was done and, if so, what type of test was done, the date and time the test was resulted, and the result of the test; the primary problem in the chart for the hospital admission; additional notes about facility policies for COVID-19 screening, testing and housing of patients as available in the medical documentation; and the chart reviewer’s determination of whether the patient should have been counted based on the NHSN definition. This information was then used to assess which reporting system (computer extraction or facility reporting) more accurately applied the NHSN definition for each facility. The Excel™ template and extensive communication between the reviewers reinforced that the data were collected consistently.

**Table S1.** Characteristics of the 36 participating VA medical facilities

|  |  |
| --- | --- |
| **Characteristic** | **Number (%)** |
| Region in United Statesa |  |
| Northeast | 7 (19.4) |
| South | 15 (41.7) |
| Midwest | 7 (19.4) |
| West | 7 (19.4) |
| Complexity levelb |  |
| 1a | 14 (38.9) |
| 1b | 4 (11.1) |
| 1c | 6 (16.7) |
| 2 | 12 (33.3) |
| Bed Sizec |  |
| <50 | 7 (19.4) |
| 51-100 | 16 (44.4) |
| 100 - 200 | 11 (30.5) |
| 200+ | 2 (5.6) |
| Presence of at least one ICUd |  |
| Yes | 34 (94.4) |
| No | 2 (5.6) |
| Average monthly patient-dayse |  |
| 400 – 999 | 9 (25.0) |
| 1000 – 1999 | 10 (27.8) |
| 2000 – 2999 | 10 (27.8) |
| 3000 – 3999 | 5 (13.9) |
| 4000 – 4999 | 1 (2.8) |
| 5000 – 5500 | 1 (2.8) |

aU.S. Census Region (<https://www.census.gov/geographies/reference-maps/2010/geo/2010-census-regions-and-divisions-of-the-united-states.html>)

bComplexity level is a designation given to all VA medical facilities to indicate level of services provided and the severity of patient conditions; the current classification system has five levels with 1a being the most complex and 3 being the least complex. [Reference: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3842006/>]

cAs reported by facilities in data collection for this project

dBased on Intensive Care Units (ICUs) on record in the VHA Inpatient Evaluation Center database for June 2020

eCalculated for calendar year 2019

**Table S2.** Facility summaries for degree of matching between computer-extracted data and facility-reported data, by total hospitalized patient counts and distinct patient comparisons.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Facility | Analysis of Daily Counts | | | | Analysis of Distinct Patient Comparisons | |
| Average daily patient count | | No. days (%) that daily counts matched | For days that counts did not match, median difference in counts (range)a | For days that daily counts matched, percent of patients matching by PIIb | For days that daily counts did not match, percent of patients matching by PIIc |
| Computer-extracted | Facility-reported |
| A | 2.0 | 1.1 | 3 (30) | 1 (1 – 3) | 100 | 55 |
| B | 0.1 | 0.2 | 9 (90) | 1d | No datae | 50 |
| C | 0.4 | 0.4 | 10 (100) | N/A | 100 (4/10)f | N/A |
| D | 1.8 | 1.9 | 7 (70) | 1 d | 100 | 39 |
| Eg | 11.0 | 12.8 | 1 (10) | 2 (1 – 5) | 80 | 76 |
| F | 4.4 | 3.8 | 4 (40) | 1 (1 – 3) | 100 | 75 |
| G | 7.2 | 7.9 | 4 (40) | 1 (1 – 2) | 100 | 85 |
| H | 0.8 | 1.0 | 8 (80) | 1 d | 100 | 33 |
| Ig | 5.7 | 9.1 | 0 (0) | 3.5 (2 – 5) | N/A | 59 |
| Jg | 12.7 | 10.6 | 0 (0) | 3 (1 – 5) | N/A | 63 |
| K | 1.6 | 1.1 | 5 (50) | 1 (1 – 2) | 100 (3/5)f | 32 |
| L | 3.4 | 3.5 | 6 (60) | 1 (1 – 2) | 100 | 50 |
| Mg | 8.1 | 8.1 | 1 (10) | 1 (1 – 4) | 44 | 34 |
| N | 0.7 | 0.6 | 9 (90) | 1 d | 100 (4/9) f | 0 |
| O | 2.4 | 2.4 | 10 (100) | N/A | 100 (9/10) f | N/A |
| P | 3.7 | 3.6 | 7 (70) | 1 d | 95 | 72 |
| Q | 4.7 | 4.2 | 5 (50) | 1 d | 91 | 80 |
| Rg | 5.5 | 1.5 | 1 (10) | 4 (2 – 9) | 100 | 24 |
| S | 1.6 | 1.1 | 4 (57) | 1 (2 – 3) | 100 | 36 |
| Tg | 8.8 | 5.2 | 0 (0) | 3 (2 – 7) | N/A | 57 |
| U | 3.5 | 3.0 | 5 (50) | 1 d | 95 | 77 |
| Vg | 1.3 | 0.5 | 1 (10) | 1 (1 – 2) | No datae | 22 |
| W | 0.9 | 0.1 | 3 (30) | 1 (1 – 2) | No datae | 0 |
| Xg | 4.8 | 1.2 | 0 (0) | 4 (1 – 6) | N/A | 22 |
| Y | 1.8 | 1.7 | 9 (90) | 1 d | 100 | 50 |
| Zg | 1.5 | 3.3 | 0 (0) | 2 (1 – 3) | N/A | 27 |
| AA | 1.7 | 1.8 | 7 (70) | 1 d | 100 | 61 |
| BB | 1.2 | 0 | 3 (30) | 2 (1 – 2) | No datae | 0 |
| CC | 1.6 | 1.0 | 4 (40) | 1 (1 – 2) | 92 (2/4) f | 24 |
| DD | 13.1 | 12.9 | 8 (80) | 1 d | 100 | 93 |
| EE | 4.3 | 2.8 | 2 (20) | 1.5 (1 – 4) | 58 | 54 |
| FF | 2.6 | 4.1 | 4 (40) | 3.5 (1 – 4) | 63 | 47 |
| GGg | 11.4 | 12.0 | 1 (10) | 1 (1 – 2) | 100 | 85 |
| HHg | 7.2 | 7.7 | 0 (0) | 2 (1 – 7) | N/A | 58 |
| IIg | 1.3 | 2.0 | 1 (10) | 1 d | 0 | 42 |
| JJ | 2.4 | 2.5 | 9 (90) | 1 d | 100 | 67 |
| Mean (S.D.) | 4.1  (3.9) | 3.8  (4.1) | 4  (3) | 1 (1 – 9)h | 95  (16) | 48  (30) |

a For only days where counts did not match between computer-extracted and facility-reported data, the difference in the count between the two sources was calculated for each such day and the median of the differences is shown for each facility with the minimum and maximum difference shown in parentheses.

bPII, personally identifying information (patient first name, last name, date of birth, and last four digits of Social Security number).

c For each facility that had one or more days of non-matching counts between the computer-extracted and facility-reported data, patient PII were reviewed for each non-matching day. Unique patients between the two data sources were reviewed for each non-matching day to determine what percentage of the patients matched.

d No range provided because there was only one day at the facility when counts did not match between the two data sources, or because all mismatching days had the same difference in counts between the two sources.

eOn all matching days, the patient count was 0.

fRatio in parenthesis indicates the number of matching days for which there were patient counts; on the remaining days, the patient count was 0.

gFacilities for which patient medical charts were reviewed to understand discrepancies in counts between the two data sources.

h These values represent the overall median and range for the 34 facilities with data in the column.

**Table S3.** Chart review assessment for selecta facilities with poor matching of counts between computer-extracted and facility-reported data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Facility** | **Which data source had more cases for each of the 10 days?** | **Observations from Chart Reviews** | **Which data source was more accurateb based on chart review?** |
| E | Variable:  7/9 days, Facility Report had more cases; 2/9 days, Computer Extract had more cases | Most patients matched between the data sources; mismatching mostly occurred because of the Facility Report or the Computer Extract counting patients without symptoms undergoing admission screening | Both were similarly accurate |
| I | Facility Report | Facility reported patients who had a past positive result but no current symptoms on admission. | Computer Extract |
| J | Computer Extract | The Extract pulled patients who did not have a COVID-19 test done or who had a negative result. | Facility Report |
| M | Variable:  4/9 days, facility report had more cases; 5/9 days, computer extract had more cases | The Extract pulled patients without symptoms but who were tested (and resulted negative) per universal testing policy; the Facility Report pulled patients admitted from a skilled nursing facility and who had a past positive test (but COVID-19 considered resolved) | Neither |
| R | Computer Extract | Facility protocol for universal SARS-CoV-2 testing on admission; Extract pulled these patients due to pending test (i.e., PUIc) even when there were no symptoms | Facility Report |
| T | Computer Extract | Extract pulled patients without symptoms and with negative test results | Facility Report |
| V | Computer Extract | The Extract pulled patients who did not have a COVID-19 test done or who had a pending test without symptoms (and negative result). | Facility Report |
| X | Computer Extract | Extract pulled patients that had symptoms consistent with COVID-19, but who tested negative; unable to determine if the patients were considered “PUI” (i.e., Suspect) by the extraction algorithm | Facility Report |
| Z | Facility Report | Facility Report included long term care residents in the hospitalized patient counts | Computer Extract |
| GG | Facility Report | Most patients matched between the data sources; mismatching mostly occurred because the Facility Report counted a patient on a day when the patient should not have been counted based on the status of the patient at the 10:00 AM time for the daily count (e.g., patient was admitted to the hospital with a COVID-19 diagnosis after 10:00 AM). The Computer Extract correctly did not count the patient until the next day. | While both were similarly accurate, the Computer Extract was more accurate when patients didn’t match. |
| HH | Variable:  First 7 days, facility report had more cases; last 3 days, computer extract had more cases | Facility had local protocols to designate patients as PUI without symptoms (e.g., transfer from skilled nursing facility); for last three days, Facility Report missed ongoing cases as having been discharged (Extract still counted them). | Computer Extract |
| II | Facility Report | Facility correctly reported suspect case with symptoms and test pending (Extract missed the case); Facility correctly reported same patients each day with severe COVID-19 illness that did not resolve during the 10-day project window (Extract stopped reporting the case once patient was removed from isolation, even though patient was still hospitalized with severe COVID-19 symptoms). | Facility Report |

a Facilities were selected for chart reviews if they had zero days or only one day of matching hospitalized patient counts when comparing computer-extracted and facility-reported data. Facility letters in the first column correspond to facility letters in Table S2.

bAccuracy was determined by assessing from chart reviews if the patients counted by each data source should have been counted based on NHSN reporting definitions.

cPUI, Person Under Investigation