

LETTERS TO THE EDITOR

RESEARCH

Accuracy of ICD-10 codes for identifying skilled nursing facility residents with lab-confirmed COVID-19

INTRODUCTION

In the absence of widely available SARS-CoV-2 testing data, administrative claims will be important for identifying COVID-19 cases among skilled nursing facility (SNF) residents in future research. However, the validity of International Classification of Diseases, Tenth Revision (ICD-10) codes for identifying residents with lab-confirmed infection has not been established. Using electronic health record (EHR) data, we examined the frequency of COVID-19 ICD-10 documentation among SNF residents with lab-confirmed SARS-CoV-2 infection.

METHODS

We used testing records from the EHR of a large multistate long-term care provider to identify SNF residents with SARS-CoV-2 infection confirmed by polymerase chain reaction or rapid antigen testing between March 16, 2020 and May 26, 2021. We queried residents' EHR problem lists for the primary ICD-10 code U07.1 (COVID-19, virus identified)¹ as well as Z86.16 (personal history of COVID-19), and J12.82 (pneumonia due to COVID-19).²

We examined variation in documentation of the U07.1 code based on whether residents were symptomatic or asymptomatic at the time of their initial positive test, as derived from change in condition notes completed by nurses.³ Residents with new symptoms from 5 days before up to 14 days after their initial positive test were considered symptomatic. We also evaluated variation in ICD-10 documentation across months. Finally, we examined the correlation between the initial date of code U07.1 and the resident's first positive test date in order to understand whether the initial date of the ICD-10 code could be used to reasonably approximate infection onset.

RESULTS

As of May 26, 2021, 17,194 unique SNF residents had lab-confirmed SARS-CoV-2 infection across 294 SNFs in

25 states. Of these, 15,949 residents (92.8%) had ICD-10 code U07.1 in their EHR problem list (Table S1). Documentation of U07.1 was slightly more common among symptomatic (7064 of 7449 residents, 94.8%) versus asymptomatic residents (8885 of 9745, 91.2%). One-third of all residents ($n = 5703$, 33.2%) had code Z86.16 listed, and this was slightly more common among asymptomatic (35.2%) versus symptomatic (30.5%) residents. Only 156 residents (0.9%) had code J12.82 code listed. Expanding the EHR query for U07.1, Z86.16, or J12.82 identified only 76 additional residents with lab-confirmed infection compared with U07.1 alone (93.2% vs 92.8%).

The U07.1 code was more often in the EHR of residents who tested positive from March to May 2020 (97.5%) or June 2020 to August 2020 (97.4%). From September 2020 to February 2021, the frequency of U07.1 among positive cases dropped to 88% (Table S2). There were 226 cases of COVID-19 from March 2020 to May 2021, of which only 138 (61.1%) had code U07.1 present. Among 15,949 residents with U07.1 in their EHR problem list, the initial date of the code was strongly correlated with the first positive test date identified from testing records ($r = 0.99$) with a mean difference of 0.16 days (Figure 1). Due to insufficient resolution dates in the problem list, it does not appear that ICD-10 codes are able to accurately capture the cases of reinfection.

DISCUSSION

To our knowledge, this is the first attempt to validate ICD-10 codes against clinical testing data for the purpose of identifying COVID-19 cases. These findings from a large multistate sample of SNF residents suggest that the U07.1 code can be used with reasonable accuracy to identify individual cases of COVID-19 in this population. Additional codes, including Z86.16 and J12.82, do not appear to meaningfully improve the sensitivity of detecting lab-confirmed cases.

Although these data come from a geographically diverse sample of SNF residents, all of the facilities are

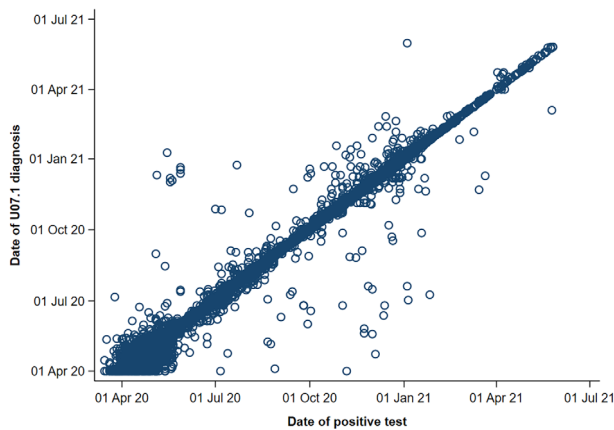


FIGURE 1 Correlation between positive SARS-CoV-2 test date and U07.1 ICD-10 code start date. Pearson's correlation coefficient (r) = 0.99

owned and operated by a single long-term care provider. Still, because patient-level testing data are not widely available to researchers, these results provide important validation to support the use of ICD-10 codes for future research using claims data. Although we queried the EHR problem list, rather than actual claims files, these are the ICD-10 codes that SNFs submit for clinical reimbursement and are expected to align well with actual claims submissions. Because code U07.1 triggers important reimbursement for SNFs, it is likely given high priority across most long-term care providers.

Symptom presentation may have contributed to the lower frequency of U07.1 being coded in Fall 2020/Winter 2021 versus Spring/Summer 2020, since a greater proportion of cases of COVID-19 were asymptomatic in the later months. Another potential explanation could be a lag in the documentation of ICD-10 codes, suggested by the observation that only 61% of cases from March 2020 to May 2021 had U07.1 listed. This temporal variation should be considered when using ICD-10 codes to compare cases at different stages of the pandemic or in settings where high numbers of asymptomatic cases are suspected.

KEYWORDS

COVID-19, long-term care nursing home, SARS-CoV-2, skilled nursing facility

ACKNOWLEDGMENTS

We thank Jeffrey Hiris from Brown University; and Richard Castor, Cliff Boyd, and Denine Hastings of Genesis HealthCare for their extensive data management support. The authors thank National Institute on Aging, 3P01AG027296-11S1 (PI: Vincent Mor) for funding.

FUNDING INFORMATION

National Institute on Aging, 3P01AG027296-11S1

CONFLICT OF INTEREST

Vincent Mor is Chair of the Scientific Advisory Board at NaviHealth, Inc., former Chair of the Independent Quality Committee at HCR ManorCare, and former Director of PointRight, Inc., where he holds less than 1% equity. The other authors have no conflicts of interest to disclose.

AUTHOR CONTRIBUTIONS

Study design: CMS, EMW, VM. *Data acquisition:* CMS, EMW, VM. *Data Analysis:* CMS, EMW. *Data Interpretation:* all authors. *Writing and critical revision of the manuscript:* all authors.

SPONSOR'S ROLE

The National Institute on Aging had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Christopher M. Santostefano BSN, RN¹

Elizabeth M. White PhD, APRN¹ 

Richard A. Feifer MD, MPH²

Vincent Mor PhD¹

¹*Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, Rhode Island, USA*


²*Genesis HealthCare, Kennett Square, Pennsylvania, USA*

Correspondence

Christopher M. Santostefano, Department of Health Services, Policy and Practice, Brown University School of Public Health, Providence, RI, USA.

Email: christopher_santostefano@brown.edu

ORCID

Elizabeth M. White  <https://orcid.org/0000-0003-4175-8662>

REFERENCES

- Centers for Disease Control and Prevention. New ICD-10-CM code for the 2019 novel coronavirus (COVID-19). April 1, 2020. Accessed May 27, 2021. <https://www.cdc.gov/nchs/data/icd/Announcement-New-ICD-code-for-coronavirus-3-18-2020.pdf>.
- Centers for Disease Control and Prevention. New ICD-10-CM code for the 2019 novel coronavirus (COVID-19). December 3, 2020.

Accessed May 27, 2021. <https://www.cdc.gov/nchs/data/icd/Announcement-New-ICD-code-for-coronavirus-19-508.pdf>.

- White EM, Santostefano CM, Feifer RA, et al. Asymptomatic and presymptomatic severe acute respiratory syndrome coronavirus 2 infection rates in a multistate sample of skilled nursing facilities. *JAMA Intern Med.* 2020;180:1709-1711.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

DOI: 10.1111/jgs.17427

COVID-19 transmission and case fatality in long-term care facilities during the epidemic first wave

INTRODUCTION

Institutionalized populations experience a higher risk of several infectious diseases especially in settings with frequent interpersonal close contact and overcrowding.¹ Besides the higher risk of infection, institutionalized people tend to be more vulnerable to its consequences as they are older and present multiple and severe comorbidities.^{1,2}

Large proportions of cases and deaths due to COVID-19 occurred in long-term care facilities (LTCF) in several countries: in May 2020, 37–66% of COVID-19 fatal cases occurred in LTCF.³ In Portugal, by November 2020 about 34% of COVID-19 deaths had occurred in these facilities.⁴

The heterogeneity of COVID-19 impact raises the hypotheses of transmission inside LTCF merely reflect the infection dynamics in the surrounding community or COVID-19 more severely affect those institutionalized. Thus, we estimated the cumulative incidence and the case fatality rate of SARS-CoV-2 infection in LTCF, considering age-adjusted comparisons to the general population.

METHODS

Eight public health units from the Portuguese Northern region provided data for the period of April to June 2020 on age, diagnosis and prognosis of 13,736 individuals (7642 residents and 6094 workers) living or working in 247 LTCF (about 10% of all LTCF in mainland Portugal, 36% of its Northern region). Data collection spanned from April to

Table S1. ICD-10 COVID-19 code presence among SARS-CoV-2-positive skilled nursing facility residents, by symptom status

Table S2. Prevalence of ICD-10 COVID-19 codes for skilled nursing facility residents with lab-confirmed SARS-CoV-2 infection, by quarter

June 2020. The number of SARS-CoV-2 infections in the Portuguese population and of COVID-19-related deaths by age was obtained from the Directorate-General of Health, and data on the population size by age groups from the Portuguese National Institute of Statistics.

The Northern Health Regional Administration Ethics Committee approved the study (CE/2020/35).

We estimated the cumulative incidence of SARS-CoV-2 infection and case fatality in each setting and in the overall sample. We compared the frequency of SARS-CoV-2 infection and death in LTCF to the community using indirect standardization. Standardized infection and mortality rates were calculated as the ratios of the incident (diagnosed) cases or deaths in LTCF residents and/or workers, and the sum of the predicted individual probabilities of infection or death from all population according to age groups (aggregated in classes of 10 years). Confidence intervals (95%) were calculated according to Hosmer and Lemeshow.⁵

RESULTS

A positive RT-PCR was observed in 416 (3.0%) individuals, 3.5% of residents, and 2.4% of workers. Incidence proportions by LTCF varied from 0.3% to 100%.

Age distribution was available for 10,531 (76.7%) individuals, 5374 (70.3%) residents (median age 85 years-old [P_{25} – P_{75} = 80–90]), and 5157 (84.6%) workers (median age 46 years-old [P_{25} – P_{75} = 36–55]). In the subsample with available data on age, we observed 166 infected individuals in LTCF while only 98 were expected according to the age incidence in the general population (Figure 1). Residents presented a 37% higher risk and