DSEN ABSTRACT
Reliability of COVID-19 case definitions in administrative and clinical databases

Summary

- In 2020, WHO created two new ICD-10 codes related to SARS-CoV2 infection
- We tested the reliability of these codes compared to PCR testing across 8 provinces in three data sources
- Using hospital "data warehouses" in Quebec and Ontario, the sensitivity and PPV of code U07.1 was relatively high. PPV estimates for code U07.1 were similar across the data warehouses, and both were higher when the first PCR test was done within 24h after admission
- In provincial administrative data, ICD-10 code U07.1 overall had high sensitivity in hospitalization data. The sensitivity of code U07.1 varied much more when looking at ED data only
- In CCEDRRN data, ICD-10 code U07.1 had high sensitivity and PPV, with better operating characteristics in those hospitalized versus those discharged from the ED. The sensitivity of ICD-10 code U07.2 was very poor.
- In sum, ICD-10 code U07.1 has high sensitivity, PPV, and specificity, particularly in hospitalization data. The sensitivity of ICD-10 code U07.2 was poor, as it was used infrequently when PCR testing is indeterminate or unavailable, and more often used for PCR-confirmed cases (where U07.1 should be used) or PCR negative scenarios (where U07.2 does not apply)

Authors: S. Bernatsky, C. Moura, C. Hohl, CAN-AIM investigative team

For more information, please contact :
sasha.bernatsky@mcgill.ca

What is the current situation?

In response to the COVID-19 pandemic, the World Health Organization (WHO) released two new International Classification of Diseases Revision 10 (ICD-10) codes to identify COVID-19 cases: U07.1 (COVID-19, virus identified: lab-confirmed SARS-CoV2 infection) and U07.2 COVID-19 (suspected/probable SARS-CoV2 infection, lab confirmation inconclusive/unavailable).

What was the aim of the study?

To investigate the operating characteristics of ICD-10 based case definitions of both confirmed and suspected/probable SARS-CoV2 infection.

How was the study conducted?

We used three data sources: i) hospital data warehouses from Quebec (**CODA-19 registry**), Ontario (the General Medicine Inpatient Initiative **GEMINI**), and The Ottawa Hospital (**TOH**); ii) the COVID-19 multi-province data accessed via Canada's Health Data Research Network (**HDRN**); and iii) the national Canadian COVID-19 Emergency Department Rapid Response Network (**CCEDRRN**). Data spanned BC, Alberta, Saskatchewan, Manitoba, Quebec, Ontario, and maritime provinces. Analyses were restricted to adults (≥ 18 years) with a polymerase chain reaction (PCR) SARS-CoV-2 test or a COVID-19 ICD-10 diagnostic code within electronic health data. Descriptive statistics were used to characterize the study populations, healthcare encounters and PCR test characteristics (e.g., test results and specimens collected). PCR results were the reference standard. Stratified analyses included demographics/timing of the first PCR.

What did the study find?

ICD-10 code U07.1 has high sensitivity, specificity, and positive predictive value (PPV) in identifying PCR-confirmed COVID-19 cases, particularly from hospitalization data but may miss cases within non-hospitalized ED patients (e.g., discharged from ED before the result of the test is known). Sensitivity, specificity, and PPV were higher if PCR testing was done within a 24–48-hour period before hospital admission/ED presentation. Results from hospitalization data were comparable across the provinces, and sensitivity and specificity tended to be higher in older individuals and urban areas. The sensitivity of ICD-10 code U07.2 for suspected COVID-19 was poor. This code seems to be used infrequently when PCR testing is indeterminate or unavailable, and more often used for PCR confirmed cases (where U07.1 should have been used) or PCR negative scenarios (where according to WHO's definition, U07.2 does not apply)

- Data warehouse sensitivity estimates for hospitalized patients were as high as 97.8% (95% confidence interval, CI 95.9-98.9, in GEMINI).
- In CCEDRRN data, the sensitivity of ICD-10 code U07.1 for lab-confirmed SARS-CoV-2 was 93.6% (95% CI 93.0-94.1) in those hospitalized from ED and 83.0% (95% CI 82.1-83.9) in those discharged from ED. The sensitivity of code U07.2 for suspected COVID-19 was very low both in people discharged from the ED and in those who were hospitalized from the ED.
- In HDRN data, ICD-10 code U07.1 sensitivity was good, and highest in Ontario for those hospitalized at 90.9% (95% CI 90.8-91.0). Manitoba had the lowest sensitivity for those discharged from ED at 39.8% (95% CI 38.0%-41.6%) compared to 71.8% (95% CI 71.7-71.9) in Ontario.