Article ID: 2023-0033

Article title: ICD hospital diagnostic codes to identify SARS-CoV-2 infections are valid in reference to PCR results

Authors: Cristiano S. Moura PhD, Autumn Neville BA, Fangming Liao MSc, Bijun Wen PhD, Fahad Razak MD MSc, Surain Roberts PhD, Amol A. Verma MD MPhil, Sasha Bernatsky MD PhD

Reviewer 1: Dena Schanzer

Affiliation: Public Health Agency of Canada, Infectious Disease and Emergency Preparedness Branch

Is it possible to check the U07.1 coding between the GEMINI (study) discharge records and the DAD? I would suspect that the coding (or miscoding) should be identical. Is it? If not, which database is more accurate, and why?

GEMINI data are comprised of administrative data (CIHI DAD, NACRS) linked to clinical data extracted from hospital electronic health records. GEMINI receives discharge diagnoses codes as reported by hospitals for the DAD. While we expect these to be essentially identical, changes could be made post-hoc by CIHI after they receive the data. There is no way to compare the data we hold to what is in the final DAD dataset at the moment because our data are not linked to CIHI data.
We have clarified this point in the text, Methods: *GEMINI receives discharge diagnosis codes as reported by hospitals.*

What was the testing protocol for COVID-19 in the hospital setting and how did this change over the study period? I seem to recall that initially, only persons with a travel history from specific high-risk countries were tested.

Was everyone admitted to the hospital eventually tested? And for which period? Did this change after the study period? At what point was testing no longer routine? If you have this information, it would be very helpful in assessing the quality of diagnostic data as a measure of disease burden.

Each hospital implemented their own testing strategy, and these could change over time. All hospitals used PCR testing. Testing was largely for patients who were symptomatic, although there was universal testing of all admitted patients at some hospitals. We have added the following statement in the text, Limitations and strengths: *Testing policies for SARS-CoV-2 varied across hospitals and fluctuated over time (Table 2), but generally, all patients with typical or atypical symptoms associated with COVID-19 were tested and, in some settings, all admitted patients were tested. Our findings are generally consistent across time periods and minor fluctuation may be explained by policy changes.*

What proportion of hospital admissions in Ontario are included in the study? Which Ontario hospitals were included in the study? Where all GEMINI hospitals included? Are these results likely generalizable to all of Ontario? Did the accuracy of coding vary by hospital? Where was the ICD-10 coding done? Does this vary by hospital across Ontario?

Public Health Ontario provided publicly available data on the number of COVID-19 hospitalizations in the form of a weekly or monthly report on the cumulative number of hospitalizations [1]. The total number of admissions for COVID-19 during the study period (Jan15-Dec 31, 2020) is N= 10,874. This means that the positive COVID-19 cases in our cohort (i.e., ICD-10 U07.1 used N=1,350) represented about 12.5% of all COVID-19 positive admissions in Ontario during that time.

We included data from seven GEMINI hospitals. GEMINI expanded from 7 Toronto and

Mississauga hospitals to 30 hospitals across Ontario in 2021-2022 (Phase 2). At the time of analysis for this project, only data from the first 7 hospitals was available (Phase 1) representing >90% of all admissions.

The following has been added to Methods: Setting and identification of subjects/eligibility:

The hospitals were in the Greater Toronto Area (Mount Sinai, Sunnybrook Hospital, St-Michael's Hospital, Toronto Western Hospital, Toronto General Hospital) and Mississauga (Trillium Health Partners Credit Valley and Mississauga Hospitals). **Our results are likely generalizable to all of Ontario because our findings are consistent with literature from the U.S. and other jurisdictions. COVID-19 was a novel disease with a lot of attention and scrutiny, so it makes sense that it's coding is very accurate.**

Thank you for your comment considering possible variation in the accuracy of coding vary by hospital. We included Table 3 which includes 3 of the seven hospitals (excluded names due to confidentiality) that had PCR data required for the analyses as described in the STARD flow diagram in the manuscript (Figure 1).

ICD-10 codes are coded manually by trained reviewers, who review the medical record at discharge and assign diagnosis codes. This process is similar at all hospitals in Ontario.

1. https://www.publichealthontario.ca/en/Data- and-Analysis/Infectious-Disease/COVID- 19-Data-Surveillance/COVID-19-Data- Tool?tab=trends

For influenza and other respiratory viruses, testing is not yet routine enough to interpret viral identification ICD-10 codes as representative of the actual disease burden. Methods to estimate excess mortality have been reported by various countries for COVID-19, and these estimates indicated that only a fraction of all deaths likely due to COVID-19 were identified as such.

Estimating excess respiratory hospitalizations attributable to COVID- 19 may be a challenge, as hospitals reduced surgeries and other procedures to have capacity for COVID cases. The long delay from symptom onset to hospitalization could also make the PCR results more difficult to interpret. On the one hand, viral load often declines with time from symptom onset. On the other hand, hospitals initially required two negative tests before considering that the case had cleared the virus.

The problem was that many people remained positive well after an apparent full recovery (30+ days), and this protocol was eventually drop. Hence it seems quite plausible that the U07.1 admissions could be either an over or under count of admissions attributable of a SARS-CoV-2 infection.

We appreciate the reviewer's insightful comment regarding the challenges associated with estimating the disease burden of respiratory viruses. We acknowledge the importance of addressing these considerations to ensure a comprehensive understanding of our findings with using ICD-10 codes for disease burden estimation.

Is the ratio of U07.1/U07.2 (i.e., lab confirmed/clinical dx) consistent over time? It would be helpful to report this ratio, as this would be an additional indicator of the limitations of interpreting U07.1 hospitalizations.

Thank you for this comment. We have included a plot to address this item (Figure 2). We plotted the proportion of U07.1 over the sum of U07.1 and U07.2, instead of the ratio U07.1/U07.2. This is because U07.2 usage is zero for some time periods, which leads to mathematical errors. Proportion of U07.1 (orange line) is shown only when the sum of U07.1 and U07.2 is greater than 1, for meaningful presentation of data.

While I would expect a relatively low error rate in the ICD-10 coding, I am rather surprised by the close

agreement between the rule-based clinical diagnoses and U07.1 admissions, and the implied level of agreement between lab confirmation and clinical diagnosis. This is rather unusual for respiratory viruses, though may possibly be explained by the stringent infection control measures (masks and lockdowns) that reduced the spread of most other viruses as well as COVID. The manuscript should elaborate on this further to ensure proper interpretation. I suggest including further evaluation work of your rule-based clinical dx: a direct comparison with lab confirmed admissions (primarily U07.1) as well as the U07.1 comparison; U07.2 admissions (COVID-19 was diagnosed clinically or epidemiologically but laboratory testing is inconclusive or not available); Z03.8 admissions (suspected but ruled out by negative laboratory results).

Would these two additional categories not be closely related to your clinical dx criteria? I would expect that a clinical dx rule (based on fever AND at least one additional feature: oxygen saturation < 93% OR oxygen supplementation OR mechanical ventilation) would be indicative of many conditions, such as any respiratory infection especially among persons with COPD or CHF and even pneumonia. If the clinical dx rule stands up further scrutiny, please provide the details of the clinical dx rule so that others interested in assessing the full disease burden can asses it and potentially make use of it.

We appreciate the reviewer's insightful comments on our study findings. We would like to clarify that, based on suggestions of another reviewer, we have decided to remove the LCA results, including the rule-based clinical diagnoses, from the revised version of the manuscript. However, we believe it is important to address the reviewer's concerns and provide additional clarifications on this topic.

We agree with the reviewer's point that the agreement between the ICD-10 code U07.1 and symptoms is expected to be low, which aligns with our observations. In our study, considering only non-missing observations, half of the cases (616 out of 1,126) assigned the U07.1 code were also classified as positive based on the diagnosis criteria. This indicates a sensitivity of the symptom-based criteria of 55% if we were to assume the ICD code as a gold standard.

Furthermore, we found that 2,342 cases had a positive symptom assigned, resulting in a potentially low positive predictive value (PPV) of 25% under the same assumption. These findings emphasize the challenges in relying solely on the ICD-10 code U07.1 or symptom-based criteria for accurate identification of confirmed SARS-CoV-2 infections. The limitations of using diagnostic codes and the inherent variability in symptom presentation contribute to the observed discrepancies. It is important to consider these factors when interpreting the validity and accuracy of the code in capturing the true disease burden.

Please explain why you found such close agreement between your clinical dx and PCR confirmation when the estimated sensitivity of PCR testing in a clinical setting is 80%. When estimating the error rates for coding, PCR sensitivity is not important, however, it does figure into the comparison between a clinical dx and PCR dx, as does the proportion of patients who were tested, and whether clinical factors affected who was tested.

We appreciate the reviewer's question. Removing the LCA results from the revised version excludes the discussion on the agreement between clinical diagnoses and PCR testing.

However, we can provide additional insights based on our previous analyses to address the reviewer's question. In our previous examination of the non-missing cases, we found that out of the 395 PCR-confirmed cases, 205 were also classified as positive based on the diagnosis criteria. This indicates a sensitivity of the symptom-based criteria of 59% if we assume the PCR test as the gold standard. Similarly, 2,013 cases had a positive symptom assigned, resulting in a PPV of 10% under the same assumption.

Furthermore, when analyzing the concordance between clinical diagnosis and PCR confirmation, we observed relatively low agreement. The estimated PPV of the symptom presentation to a

positive PCR test was only 10%, indicating that only a small proportion of patients with symptoms actually tested positive for SARS-CoV-2. These revised analyses clearly highlight the limitations of relying solely on clinical diagnosis in accurately capturing the true prevalence of COVID-19 compared to PCR testing.

Although these findings are not included in the revised version, we believe they are important to consider in the context of our study and the validity of the ICD-10 code U07.1. We appreciate the opportunity to provide these additional insights.

Reviewer 2: Giulio DiDiodato

Affiliation: Royal Victoria Regional Health Centre, Critical Care Medicine

The authors have conducted an important study on the diagnostic accuracy of routinely collected data.

Some comments/questions

1) My only concern with this study was the inclusion of the LCM. There are several issues which need clarification.

a) I would like to better understand the rationale for including the LCM analysis - as advances in diagnostic testing with RAT have occurred since December 2020, the need for this approach to the diagnosis of COVID-19 disease seems to be irrelevant, especially given the test characteristics of PCR, RAT and serological testing. One, I suppose, could imagine an LCM that incorporated all these diagnostic tests, along with the clinical criteria, as an approach to the diagnosis of COVID-19 but I don't see how this would provide a meaningful improvement in the diagnostic accuracy of the U07.1 code.

b) In the LCM analysis conducted in this study, the authors combined the PCR result, the discharge diagnostic code U07.1 and clinical criteria as indicator variables to assess their hypothetical 2-class LCM. While the clinical criteria indicator could be considered conditionally independent of the other 2 criteria, the PCR result and the U07.1 indicator variables are almost perfectly correlated as per their previous assessment. In addition, U07.1 coding would not be considered a 'diagnostic' test since it is based on the presence or absence of a positive or negative PCR result, respectively. This would make these indicators conditionally dependent which violates one of the major assumptions of the LCM approach, resulting in only 2 indicator variables being included in the assessment of the hypothesized 2-class LCM.

c) The authors appear to have tested only a 2-class LCM, as opposed to testing and comparing different x-class models to determine which would provide the best fit. The rationale for this is not provided. It could be that a 3-class LCM that contained an 'indeterminate' group might be better than the hypothesized 2-class model. If this were the case it would be difficult to demonstrate since with only 2 indicator variables, the degrees of freedom = 0 for the 3-class LCM which could create problems with model stability and prediction of confidence/credible intervals.

d) given all these issues, i am not sure that the paper benefits from inclusion of the LCM as it is, and it might be best to consider removing it altogether. If the authors wish to include it, i believe there would need to be a considerable revision made to the study to further explain and describe the LCM approach and how the authors verified that the 2-class LCM was the best model.

Thank you for your thoughtful comments. Upon further consideration, we acknowledge the limitations of our approach and have elected to remove LCA from the manuscript.