Article details: 2022-0152
Title: Validity of COVID-19 diagnoses in Canadian administrative health data: a multi-province, population-based study
Authors: Lisa M. Lix PhD, Christel Renoux MD PhD, Carolina Moriello MSc, Ko Long Choi MSc, Colin Dormuth PhD, Anat Fisher, Matthew Dahl BSc, Fangyun Wu MSc, Ayesha Asaf MPH, J. Michael Paterson MSc, for the Canadian Network for Observational Drug Effect Studies (CNODES) Investigators

Reviewer 1: Yaping Jin
Institution: Department of Ophthalmology and Vision Sciences, University of Toronto
General comments (author response in bold)

Comment #1: This is a carefully designed, well-written and timely needed manuscript. This reviewer strongly recommends its publication in CMAJ Open with the following minor revisions.
We thank the reviewer for this positive response to our manuscript.

Comment #2: Abstract, “Specificity and NPV of COVID-19 diagnoses were consistently high.” Is it possible to provide a number such as >xx% after “high” so that readers will know what is the level of “high”?
Response: We have added additional information to the abstract about specificity and NPV.

Comment #3: The authors have the unique opportunity to access data from hospital inpatient discharges, ED visits and outpatient visits. Is it possible to analyze the validity of COVID-19 diagnoses by physician specialty? If diagnosed by multiple physician specialties, keep the first specialty? This information may enrich the results of the study and widen the application/citation of the publication.
Because ED records and hospital discharge abstracts are coded by health records personnel using PCR lab test data to identify the presence of COVID-19, physician specialty is less relevant in these settings. For the outpatient cohort, a sub-analysis by physician specialty would be challenging given the very large number of certified specialties and the small percentage of physician encounters accompanied by a PCR test result (2.7% in Ontario). Also, the majority of encounters are with family physicians. Therefore, we have not added this analysis to the manuscript.

Reviewer 2: Ryan Strum
Institution: HEI, McMaster University
General comments (author response in bold)

Comment #1: I don’t believe regarding this study as a validation article is the most appropriate approach, as I don’t see formal face or content validation described, and no comparisons between this measure to an accepted standard. To me, this article tests the accuracy of Covid codification in administrative databases to true positive tests in PCR. Up until the interpretation, only validation is mentioned but, in the interpretation, only accuracy is mentioned – and I think accuracy is far more appropriate here. A plausible subsequent study could evaluate the validity of these codes as a standard, but wasn’t specifically done in the methodology. I will leave this to the editors to determine the suitability of the language around these terms, but I think should be revised throughout to accuracy.

Please see our response to Comment #1 from the Editor-in-Chief.

Comment #2: The methods section needs more detail, specifically the data bases used to extract the data in the study. I would recommend inputting a subsection 'Data Source' and describing the databases individually, though briefly. I would suggest making a table to display this: rows of each province and columns that clearly show the important parts of the methods such as: inpatient data source, ED data source, outpatient data source, codification to identify Covid infections, codification used in the province, how positive tests were identified/recorded, etc. This would be a great summary to inform future research.
Thanks for this suggestion. Given the manuscript word limit, it was not possible to add information about the individual databases for each province within the manuscript. However, we have added information about the databases in Table S1 and information about the codes used to ascertain COVID-19 cases in administrative data in Table S2.

Comment #3: How was the coding of Covid in ICD-8 and ICD-9 for physician service claims identified? If the Covid codes were added to ICD-10, we know what to look for to identify Covid in ICD-10. But in the older ICD-8/9 versions, was any code added to these as well to identify Covid? If not, state how you identified Covid in this old coding (should be a part of the table I suggested above).
Please see our response to Comment #10 from the Editor-in-Chief.

Comment #4: Since you are linking the databases, please report the percentage of data linkages for each database. I don't think this needs to be in a table, but certainly important to state percentage of linkage. If there were linkage problems between datasets, should be reported as a limitation of the study
There were no linkage problems amongst datasets used in this study. As noted on page 5:
All data sources were deterministically linked at the individual level using unique anonymized health insurance numbers; only individuals with valid health insurance number were included from each province.

Comment #5: Is there any possibility of patients being in multiple sub-cohorts? If so this should be addressed, as can overinflated your analysis. Ie, patients that has a PCR test in the ED and coded as Covid, and was admitted for an inpatient as Covid – would this count as an ED visit or inpatient visit? Or both? This needs to be elaborated on to avoid double-counting a single PCR test in multiple sub-cohorts.
Please see our response to Comment #6 from the Editor-in-Chief.

Comment #6: There is no stated range in the methods of what an acceptable sensitivity is in the analysis plan, nor any citations to support this. In the interpretation, only states that inpatient was a good test result, and ED moderate. You should prespecify this in the analysis plan of the methods, and support with a citation. What is considered good in the literature? In suspicious the ED result of 60% may not be considered as moderate in the literature, and actually poor - the codification includes many patients marked as Covid but also many false negatives.
Please see our response to Comment #4 from the Editor-in-Chief.

Comment #7: Include in methods you used the STROBE reporting guideline, with citation.

We used the STROBE reporting guidelines. This information (and the accompanying reference) is noted on page 5.

Comment #8: Age of table 1 is broken into 3 groups, with one being <65 years. Given the means of each sub-cohort in Table 1 range from 43-56, I would like to get a better understanding of the spread within this really large age category. Consider breaking down into smaller groups, possibly 10-15 years each
We apologize but we did not capture smaller age groups in our study in part because of small numbers when we stratified the validity estimates by age group.

Comment #9: You should define income quintile and urban/rural. How were these determined? Administrative databases I know typically have these already pre-set, but should be stated in methods
Thanks for this suggestion. We note the following on page 7:
Area-level income quintile was based on postal code of residence and area-level household income information from the Statistics Canada Census.15,16
Rural/urban residence was based on each province's definition; for example, Manitoba rural residents lived outside of the two major urban centres of Winnipeg and Brandon.

Comment #10: Optional, add a citation to support why you chose these date ranges in the sensitivity analysis. They seemed reasonable regarding Covid
The time windows used to ascertain cases in our sensitivity analyses were based in part on previous research (Kluberg et al. 2021; Wu et al. 2022) which is noted on page 8. The larger case ascertainment windows selected for these additional analyses were expected to result in larger estimates of sensitivity.

Comment #11: First words of the results could be removed, was already stated: "After applying the study entry criteria"
Thank you for the comment. We have made the change as recommended.

Comment #12: Same for interpretation, I would remove first sentence and get right to what your study found, "Our multi-province study of the accuracy of COVID-19 diagnoses in inpatient, ED and outpatient records occurred during the first year of the pandemic when SARS-CoV-2 PCR testing was broadly promoted and openly accessible to symptomatic individuals in all three study provinces"
We have removed the first sentence of the interpretation section as recommended by the reviewer.

Comment #13: The comparison of the results to the literature seems fair for inpatient sensitivity, but not for ED or outpatient. Unless I'm reading this wrong, the literature cites percentages of >85%, but ED was 60% and outpatient 20%. I would change this statement to say the literature agreed with inpatient, but does not for ED and outpatient – and provide a reason why (literature hasn't researched ED or outpatient yet, you used population-level data instead of a sample, etc)
Thanks for this feedback. The interpretation section has been substantially revised. We focus on the range of estimates for inpatient, ED, and outpatient data and note the following on page 10:
Our multi-province validation study revealed that diagnoses for COVID-19 had PPV estimates that ranged from 17.8% to 75.0% in inpatient records, 1.0% to

81.3% in ED records, and 1.0% to 31.1% in outpatient records, illustrating the substantial variation by province and over time.
We note where our results for inpatient data are consistent with the literature and where they are not. We also note the comparison with one previous study that examined validity of ED data. We have not identified any previous studies that have examined validity of COVID-19 diagnoses in outpatient data.

Comment #14: Limitations – you state generalizability of your study to outpatients is unknown, but it is known, you just determined this in this study. I would consider removing this statement, not a limitation but a result.
Thanks for pointing out that our meaning was not clear. We have revised the sentence to clarify that generalizability of our findings about outpatient data to other provinces and territories is unknown. Specifically, we note on page 12: The generalizability of our findings to outpatient physician service claims from other provinces and territories is unknown; each province and territory implemented its own COVID-19 coding for outpatient claims, with varying directions.

Comment #15: In the interpretation, would be good to read what you think inpatient administrative data did correctly, and how ED and outpatient data can learn from this work. This would be the final message of the article I'd think.
Thanks for this helpful comment. We conclude the manuscript on page 13, noting: In summary, we identified variations in the validity of COVID-19 diagnoses recorded in different healthcare settings, geographic areas, and over time. The overall performance of diagnosis codes for COVID-19 case ascertainment was better for inpatient and ED administrative data than for outpatient administrative data in the first year of the pandemic, likely due to the greater standardization in diagnosis coding practices in the former data sources. However, over this period of time and subsequently there were changes in guidelines and testing practices that may influence the generalizability of the study findings. Nevertheless, this study provides valuable insights about the validity of administrative data sources for COVID-19 case ascertainment that can benefit population-based research and surveillance.

Reviewer 3: Jasmine Pawa
Institution: Dalla Lana School of Public Health, University of Toronto
General comments (author response in bold)

Comment #1: Page 5 - appreciate the clarity about the CNODES network and request of Health Canada.
Thank you for this positive comment.

Comment #2: It is helpful that outpatient, emergency, and hospital were all looked at (as opposed to emergency or hospital alone).
Thank you for this positive comment.

Comment #3: The methods were not completely clear. Was the laboratory data linked to other 4 administrative health databases by individual or were the quarterly rates of COVID cases being compared? This could be more explicit.

All databases were linked at the individual level. We have added a sentence in the methods section on page 5:
All data sources were deterministically linked at the individual level using unique anonymized health insurance numbers; only individuals with valid health insurance number were included from each province.

Comment #4: What are the possible explanations for the discrepancy / discordance between the admin data and the laboratory results? Did the authors have an opportunity to connect with clinicians using the codes in hospitals? Further explanation would be helpful.
There are a number of possible explanations for the performance of administrative health data. On page 13 we have concluded with the following:
In summary, we identified variations in the validity of COVID-19 diagnoses recorded in different healthcare settings, geographic areas, and over time. The overall performance of diagnosis codes for COVID-19 case ascertainment was better for inpatient and ED administrative data than for outpatient administrative data in the first year of the pandemic, likely due to the greater standardization in diagnosis coding practices in the former data sources. However, over this period of time and subsequently there were changes in guidelines and testing practices that may influence the generalizability of the study findings. Nevertheless, this study provides valuable insights about the validity of administrative data sources for COVID-19 case ascertainment that can benefit population-based research and surveillance.

Comment #5: Several limitations could be more clearly articulated:
Not accounting for rural / remote parts of the country could be stated more clearly. In the territories and Atlantic provinces, for example, it is likely that the systems in place / findings for a similar type of work would be very different.
This is an important point to be made about the validity of outpatient physician claims, for which there is greater heterogeneity in coding practices than for inpatient data. On page 12, we note:
The generalizability of our findings to outpatient physician service claims from other provinces and territories is unknown; each province and territory implemented its own COVID-19 coding for outpatient claims, with varying directions.

Comment #6: Done at a time when PCR testing was available to the general population. How can we interpret for now given the significant, as expected, shifts in pandemic response.
The evolving availability of PCR testing during the pandemic may have an impact on the generalizability of the study findings. On page 13, we note:
However, over this period of time and subsequently there were changes in guidelines and testing practices that may influence the generalizability of the study findings. Nevertheless, this study provides valuable insights about the validity of administrative data sources for COVID-19 case ascertainment that can benefit population-based research and surveillance.

Comment #7: Overall the paper could be improved by including more Canadian literature on the use of administrative databases for these types of studies, even beyond COVID-19.

We thank the reviewer for this comment. We had limited space to address this comment, but have added the following to the introduction section on page 4: There is a substantial history of conducting diagnostic validation studies for administrative health data in Canada to inform studies about the validity of COVID-19 diagnoses.[6,7]

Comment #8: As expected given the data used, the study focuses on acute care as does much of the health system activity. Further acknowledgement of the importance of community and at-home care could be considered, including options to improve data and also how all this connects with prevention (particularly given the wide-ranging impacts of the pandemic).
We note the following on page 11:
Our study provides important information about the validity of COVID-19 diagnosis coding in ED records and outpatient physician claims; the latter are particularly important for investigating the delivery of preventative care.

Comment #8: Page 10 - Significant concerns with the statement that "our findings for inpatient diagnosis coding are generally consistent with those reported in prior studies." This study reports inpatient PPV from approximately 50-65%. The comparator studies identify a PPV of approximately 80-92%. Unless I've misunderstood something in the results, this does not seem generally consistent. Similar for the sensitivity (65-85% as compared to > 95%).
Thanks for the feedback. We have significantly revised the interpretation of results on page 11:
Using US inpatient data from May to October 2020, Kluberg et al.1,2 reported sensitivity estimates of 95% and PPV of 81% for ICD-10 U07.1. Similarly, Kadri et al.4 reported sensitivity estimates of 98% and PPV of 92% from April to May 2020. Ontario estimates were consistent with these US estimates for all quarters. These US estimates were comparable to British Columbia and Manitoba estimates from the last two quarters of the study year. Thus, in the first year of the pandemic hospitals were coding COVID-19 diagnoses in inpatient data with reasonable validity, although there was variation by quarter and site.

Comment #9: More nuance in the reporting of the results and interpretation/conclusion would be helpful. If I've understood correctly, for Manitoba emergency data (page 9), for example, a sensitivity of 11% suggests with the use of this methods that of 100 positive results, roughly 10 were true positives and the rest were false negatives. Similarly, the PPV of 40% suggests that of 100 positive results, only 40 were true positives and the rest were false positives. Given the importance of the questions & potential application of this methods by others, the interpretation/conclusion would need to reflect some of these limitations.
Please see our response to Comment #4 from the Editor-in-Chief.

Comment #10: The paper makes statements such as the sensitivity is generally high which does not seem fully supported by the results.
Thanks for this comment. Please see our response to Comment #4 from the Editor-in-Chief. We have substantially revised the wording we use to describe our results.

Comment #11: This leads to some disconnect with the conclusion and perhaps too much of a stretch, given the results, in the abstract/interpretation/conclusion. It seems much of

the admin data could not be used for surveillance purposes and this should be clearly stated.

Thanks for this feedback. We now conclude the manuscript on page 13, noting: In summary, we identified variations in the validity of COVID-19 diagnoses recorded in different healthcare settings, geographic areas, and over time. The overall performance of diagnosis codes for COVID-19 case ascertainment was better for inpatient and ED administrative data than for outpatient administrative data in the first year of the pandemic, likely due to the greater standardization in diagnosis coding practices in the former data sources. However, over this period of time and subsequently there were changes in guidelines and testing practices that may influence the generalizability of the study findings. Nevertheless, this study provides valuable insights about the validity of administrative data sources for COVID-19 case ascertainment that can benefit population-based research and surveillance.