Article ID: 2020-0213

Title: Heterogeneity in testing, diagnosis, and outcome of COVID-19 across outbreak settings in the Greater Toronto Area, Canada: an observational study

Authors: Linwei Wang MSc, Huiting Ma MSc, Kristy C.Y. Yiu MSc, Andrew Calzavara MSc, David Landsman BSc, Linh Luong MPH, Adrienne K. Chan MD MPH, Rafal Kustra PhD, Jeffrey C. Kwong MD MSc, Marie-Claude Boily PhD, Stephen Hwang MD MPH, Sharon Straus MD MSc, Stefan D. Baral MD, Sharmistha Mishra MD PhD

Reviewer 1: Dena Schanzer, Public Health Agency of Canada, Infectious Disease and Emergency Preparedness Branch, Ottawa, Ont.

Comments and author response

We thank the reviewer for the constructive feedback below. As the comments and our responses to comments 1) to 4) show certain level of overlap, we have therefore compiled our responses to comments 1) to 4) into one, included after comment 1).

Comments to the Author

1) Unfortunately, I have struggled to identify the main research question. The concluding statement in the abstract suggests the authors wished to assess transmission risks in LTC facilities and shelters compared to the general population. While the authors presented time series of testing rates (per capita) positivity rates (per test) and mortality rates (per lab confirmed case), by these 3 categories, reporting delays and the constantly changing testing criteria over the study period were not accounted for. As a result, any generalization about the differences in the underlying transmission rates becomes highly speculative. For example, the main finding that the age-adjusted CFR was 1.4 times higher among LTCH residents compared with the rest of the population has a high risk of bias, as there were large differences in testing policy for LTC residents (at one point all residents were tested) and for their age-peers living in the community who were only tested if they developed severe complications.

We have edited our title and abstract accordingly, to restrict the use of the word 'risk' where not appropriate (replaced by the descriptive term 'diagnosis'), and edited our objectives for further clarity on the last paragraph of the Introduction section on page 3:

"**Title:** Heterogeneity in testing, ~~risk~~ **diagnosis**, and outcome of COVID-19 across outbreak settings in the Greater Toronto Area, Canada: an observational study"

"**Background:** We compared the testing for, ~~risk~~ **diagnosis** of, and death following COVID-19 infection across three settings (long-term care homes (LTCH), shelters, the rest of the population) in the Greater Toronto Area (GTA), Canada."

"Lessons from past epidemics suggest that disproportionate risks across settings contribute to the spread and outcomes of infection (15). Thus, a key feature of an epidemic response is quantifying heterogeneity in 'what has happened' ~~with respect to disproportionate risk~~, a process often referred to as an epidemic appraisal (16, 17). **As a first step to support epidemic appraisal**, we aimed to characterize, using the best available data sources, **patterns over time** in testing (proportion tested), diagnoses (diagnosed cases per capita, testing positivity), and outcome (death) following COVID-19 diagnosis in the GTA across three settings **for which we have data on the population size**: LTCH residents, persons using shelters, and the rest of the population."

In response to reviewers' comments, we have updated and restricted our analyses to confirmed cases reported up to May 20, 2020 instead of May 25 (as used in the original analyses) to address the impact from delays in reporting and lost to follow up.

We have updated results throughout the manuscript to reflect the change in study period. The main findings and interpretations largely remained the same.

We have added the following subsection to the Methods section on page 5:

"Study period and outcomes

iPHIS data obtained through May 31, 2020 (data-cut date) were used in our analyses, for outcomes including diagnosed cases per capita, and case fatality proportions by date of case report. OLIS data obtained through May 27 were used in our analyses, for outcomes including proportion of individuals who were tested, and the proportion of individuals tested positive by date of testing.

We defined our study period as confirmed cases reported from Jan 23, 2020 to May 20, 2020, representing approximately four months since the first confirmed case in the GTA (2). However, we used follow-up data up to May 31, 2020 to minimize potential biases from delays in completing outcomes and reporting. By the end of follow-up (May 31, 2020), less than 5% (4.3%) of confirmed cases had an unknown outcome (neither died nor resolved, influence of lost to follow-up shown in Appendix-2 Table 1). Complete entry of confirmed cases into iPHIS for a given case report date occur within 3 days (35-37), and thus we assumed complete entry by May 31, 2020 of all cases reported by May 20, 2020. We used the May 27, 2020 OLIS data-cut to analyze results of tests sampled by May 20, 2020, because 95% of laboratory results were finalized and reported into OLIS within 6 days of a given testing date (22).*"***

In addition, we performed an analysis using the latest iPHIS data available as of August 6 at the time of revision, to compare the difference in our estimates of case fatality proportion for cases reported up to May 20 using near real-time data at the time of initial analysis (May 31 iPHIS data; with less than 5% missing in outcome) vs. with near-complete follow-up data (Aug 6 iPHIS data), and found a relatively small difference in the estimates as shown in Appendix Table 2.

Appendix-2 Table-1. Comparing the estimated case fatality proportion for confirmed cases reported up to May 20, 2020 in the Greater Toronto Area using iPHIS data extracted on May 31 vs. on Aug 6, 2020.

	May 31, 2020 data cut	Aug 6, 2020 data cut
%death (case fatality proportion)	1437 (8.7%)	1672 (10.1%)
%resolved	14342 (87%)	14839 (89.8%)
%outcome unknown	711 (4.3%)	10 (0.1%)
Time from case report date to death	Min = -52 IQR = 2 – 12 Median = 7 Max = 50	Min = -29 IQR = 3 - 13 Median = 7 Max = 89

Abbreviation: iPHIS (integrated public health information system); IQR (inter-quartile range)

We have added the following sentence to the limitation section of the Discussion on page 10:
"Fifth, our case fatality estimates could be underestimated as 4.3% of cases had an unknown outcome by the end of follow-up."

We have added the following footnote to Table 1:
"[i]A total of 4.3% of confirmed cases had an unknown outcome by the end of follow-up and were considered alive in our calculation. Thus, our estimates could underestimate the case fatality proportion."

Finally, we edited the 3rd and 4th paragraph of the Discussion section on page 9-10, to <u>describe the changing testing criteria over time and to discuss the direction of bias that testing criteria may have on our estimates of test positivity rate ratio, and case fatality rate ratio between LTCH residents and the rest of the population</u>:

"The size and trajectory in per-capita diagnoses among LTCH residents and among persons using shelters likely reflect underlying differences in testing and differential risk. **Early testing in Ontario focused on symptomatic individuals with travel history, or had close contact with a confirmed case (51, 52). By March 27, symptomatic individuals in several risk groups, including residents of LTCH, were prioritized for testing (53), and after April 15, included shelters (54). LTCH testing was further expanded on April 8 to include asymptomatic individuals with potential exposures (close contacts) or in shared-rooms with a symptomatic resident (55). The changing testing criteria corresponded to the observed patterns of surge in cases identified in LTCH and shelters in April. After April 21, Ontario began to proactively test every (including asymptomatic) resident and staff in the LTCH (56, 57), which may partially explain the subsequent decline in LTCH residents' test positivity proportions.**

The **2.4**-fold higher cumulative test positivity among LTCH residents after adjustment for age and sex, **despite wider scope of testing**, suggests higher risk transmission environments; **and may actually be an under-estimate of test positivity difference**. Testing criteria outside the context of congregate settings were more risk-based (symptoms, epidemiological link or close contact/exposures) during our

study period (61-63). **Thus, if risk-based testing yields higher test positivity proportion than population-based testing, and if everyone had been tested in both groups, we would have expected an even higher test positivity rate ratio among LTCH residents versus the rest of the population. Similarly, the wider scope of testing for LTCH residents could lead to larger proportion of diagnoses of infected cases. Therefore, the infection fatality rate ratio may be even higher than the 1.4 times case fatality rate ratio observed in the current study between LTCH residents versus the rest of the population.** The higher age- and sex-adjusted case fatality rate among LTCH residents as compared with the rest of the population, may reflect underlying differences in comorbidities associated with COVID-attributable mortality and/or goals of care (64). **Future studies including information on comorbidities could help identify causal pathways between residing in LTCH and increased case fatality rate**."

2) While most public health agencies report the case fatality ratio (cumulative deaths/cumulative cases) in the regular near-real time surveillance reports, a footnote is usually included to indicate the potential for bias due to changes in testing criteria, reporting delays and survival time. While CFR is common in a surveillance report, survival analysis should be used in an observational study (as all active cases are essentially lost to follow up as of the date of extract). (For example, the CFR for LTC residents of 26% as of May 25, while a month later it is 31% and it was in the teens in early April. This is very typical of changes to the crude CFR based on near real time data.)

Please see our response following comment 1.

3) The significance of choosing a study period ending on May 25 has not been documented.

Please see our response following comment 1.

4) To avoid misinterpretation due to reporting delays surveillance reports will typically use one of the following three options: 1) report cases/deaths by date of report rather than date of dx or death; 2) wait until the outbreak/epidemic is over and all reports are in to report cases or deaths by date of symptom onset, or date of dx or date of death, or 3) identify the reporting delay. Ontario Public Health has used a form of option 3 on their web site, however, they did not estimate the reporting delay (for example, proportion of deaths reported within x days). They greyed out the period most affected by the delay in reporting.

Please see our response following comment 1.

5) The data is not well described. Dates should be defined. Each date should be identified. Is it the date reported, date of symptom onset, date of specimen procurement, date of death? The date that the data was extracted should be documented and it is usually different than the end of the study period.

We have added the following subsection to the Methods section on page 5 including the definition of date:

"iPHIS data obtained through May 31, 2020 (data-cut date) were used in our analyses, for outcomes including diagnosed cases per capita, and case fatality proportions **by date of case report**. OLIS data obtained through May 27 were used in our analyses, for outcomes including proportion of individuals who were tested, and the proportion of individuals tested positive **by date of testing.**

We defined our study period as confirmed cases reported from Jan 23, 2020 to May 20, 2020, representing approximately four months since the first confirmed case in the GTA (2). However, we used follow-up data up to May 31, 2020 to minimize potential biases from delays in completing outcomes and reporting."

We have also noted in the figure legends the definition of the date used in each figure (X-axis).

6) In light the constantly changing testing criteria, I struggled to understand the usage of the term 'risk'. As I read through the manuscript, I kept looking for the details on how transmission risk was inferred from the surveillance data provided. After reading through the manuscript a few times, I will assume that the word 'risk' should be replaced with 'proportion'.

We have edited our title and abstract accordingly, to restrict the use of the word 'risk' where not appropriate (replaced by the descriptive term 'diagnosis'), and edited our objectives for further clarity on the last paragraph of the Introduction section on page 3:

"**Title:** Heterogeneity in testing, ~~risk~~ **diagnosis**, and outcome of COVID-19 across outbreak settings in the Greater Toronto Area, Canada: an observational study"

3

*"**Background:** We compared the testing for, ~~risk~~ **diagnosis** of, and death following COVID-19 infection across three settings (long-term care homes (LTCH), shelters, the rest of the population) in the Greater Toronto Area (GTA), Canada."*

"Lessons from past epidemics suggest that disproportionate risks across settings contribute to the spread and outcomes of infection (15). Thus, a key feature of an epidemic response is quantifying heterogeneity in 'what has happened' **with respect to disproportionate risk**, a process often referred to as an epidemic appraisal (16, 17). **As a first step to support epidemic appraisal**, we aimed to characterize, using the best available data sources, **patterns over time** in testing (proportion tested), diagnoses (diagnosed cases per capita, testing positivity), and outcome (death) following COVID-19 diagnosis in the GTA across three settings **for which we have data on the population size**: LTCH residents, persons using shelters, and the rest of the population."

7) What do you mean by cross checking with data from the LTCH tracker? The cross-check process should be described in more detail. The differences, as presented in the appendices, look like the reporting delays are severe enough to delay the analysis until the data set is more complete and stable. Only the method used for cross-checking should be described in the methods section. Results of the cross-check should be moved to the results section. Actions taken as a result of this cross-check should be described and included in the discussion section. Reasons for any discrepancy should be included in the discussion section. I am also curious why the two data sets differed so much. Will this be corrected in the future? If the problem is a reporting delay, I'd suggest re-extracting your study data, once the data for your study period is considered complete.

We have included the following sentences in the 1st paragraph of the Methods section Analyses subsection on page 5 to describe the details on data cross-check:

"First, **to examine the completeness of testing data and the accuracy of classification by setting in OLIS, we compared the cumulative cases overall and by setting between OLIS and iPHIS.**"

We have moved the results of this cross-check to the 1st paragraph of the Results section on page 7, and edited for clarity:

"**Data quality**

OLIS captured/identified 92%, 97%, and 24% of confirmed cases in all settings combined, LTCH residents, and persons using shelters, respectively, compared to iPHIS (Appendix-2 Figure-1A-C). Given low sensitivity of OLIS data in identifying persons using shelters, we did not report results on testing for persons using shelters."

We have removed the comparison of iPHIS against the external dataset (LTCH tracker) with regard to cumulative diagnosed cases and death from our main text (in 2nd paragraph of the Methods section, and 1st paragraph of the Results section) and Appendix-2 Figure-1. Indeed, the provincial LTCH tracker tracks daily census of cases, thus could not capture cases which have been resolved, resulting in an underestimate of the cumulative number of diagnosed cases. The difference in the number of death per iPHIS vs. the LTCH tracker could reflect a combination of following: 1) reporting delay; 2) misclassification of case by setting; 3) difference in the definition of the GTA boundary.

To more directly examine the potential reporting delay and lost to follow-up in terms of death counts, we performed an analysis using the latest iPHIS data available as of August 6 at the time of revision, to compare the difference in our estimates of case fatality proportion for cases reported up to May 20 using near real-time data at the time of initial analysis (May 31 iPHIS data; with less than 5% missing in outcome) vs. with near-complete follow-up data (Aug 6 iPHIS data), and found a relatively small difference in the estimates as shown in Appendix Table 2.

Appendix-2 Table-1. Comparing the estimated case fatality proportion for confirmed cases reported up to May 20, 2020 in the Greater Toronto Area using iPHIS data extracted on May 31 vs. on Aug 6, 2020.

	May 31, 2020 data cut	Aug 6, 2020 data cut
%death (case fatality proportion)	1437 (8.7%)	1672 (10.1%)
%resolved	14342 (87%)	14839 (89.8%)
%outcome unknown	711 (4.3%)	10 (0.1%)
Time from case report date to death	Min = -52 IQR = 2 – 12	Min = -29 IQR = 3 - 13

	Median = 7 Max = 50	Median = 7 Max = 89

Abbreviation: iPHIS (integrated public health information system); IQR (inter-quartile range)

Statistical analyses:

8) Which statistical procedure did you use to 'examine differences' in the three settings?

We have added the following sentence as a footnote to the Table 1:

"**[i]Partial Wald tests (78) were performed to compare the difference in case fatality rate and test positivity rate across settings estimated by the quasi-Poisson regression models.**"

9) Calculations and other procedures that are not statistical procedures should be moved from the subsection on 'statistical analyses'.

We have renamed the subsection heading from 'Statistical analyses' to 'Analyses'.

Results:

10) Tables and figures corresponding to the main results should be presented in the main manuscript rather than the appendices.

We have included tables and figures corresponding to the main results as main tables and figures.

Discussion:

11) The main limitations, reporting delays and variation in testing criteria and test availability over the study period are not mentioned.

Please see our response following comment 1.

Reviewer 2: Name withheld

Comments to the author and author response

This manuscript has data that could be valuable in helping readers understand the nature of the pandemic. The authors describe the differences in how the outbreak has unfolded across different settings. They do a good job in identifying existing data and presenting the information both in writing and visually. They also appropriately identify limitations.

I do have some concerns however.

1. My major issue is with the basic concept of the paper. It seeks to describe the heterogeneity of the outbreak however, it really only looks at specific settings that appear to be purposefully selected. I believe a reframing to describe that the findings of heterogeneity only apply to these settings would be more important - i.e. the outbreak unfolded differently in long-term care homes and shelters than in the rest of the population. To truly comment on heterogeneity they would have needed to conduct a random sample of different settings across the population.

Following the reviewer's suggestion to reframe the concept of the study, we have edited the Methods section of the abstract and the objective statement on the last paragraph of the Introduction section on page 3 for clarity:

"**Methods:** We sourced person-level data from COVID-19 surveillance and reporting systems in Ontario. We calculated cumulatively, the diagnosed cases per capita, proportion tested for COVID-19, daily and cumulative positivity, and case fatality proportion **for settings which we have data on the population size** (LTCH residents, shelters, and the rest of the population)."

"Lessons learned from past epidemics suggest that disproportionate risks across settings contribute to the spread and outcomes of infection (15). Thus, a key feature of an epidemic response is quantifying heterogeneity in 'what has happened' with respect to disproportionate risks, a process often referred to as an epidemic appraisal (16, 17). **As a first step to support epidemic appraisal, we aimed to characterize,** using the best available data sources, **heterogeneous patterns over time** in testing (proportion tested), risk of diagnoses (diagnosed cases per capita, testing positivity), and outcome (death) following COVID-19 diagnosis in the GTA across three settings **for which we have data on the population size:** LTCH residents, persons using shelters, and the rest of the population."

2. Also, I am not sure that the conclusions apply to both shelters and LTC. While it is true that there were a disproportionate number of cases identified in shelters the authors do point out that the case positivity rate in shelters was less than in the general population (0.8) and the case fatality rate was also lower (0.7). Thus the higher detection in shelters may have simply been a consequence of increased testing in this setting.

Because the OLIS data were incomplete with respect to persons who use shelters or homelessness, we have removed the results pertaining to testing (proportion of individuals who were tested, and the proportion of individuals who tested positive) for persons using shelters from our main text, Table1, and Figure 3. Instead, we added the following 'footnote' to Table 1:
"**ᶠWe did not show results pertaining to testing for persons using shelters due to low sensitivity in identifying persons using shelters in the Ontario Laboratory Information System (OLIS) testing data. A total of 1398 individuals had an indication of 'homelessness' in OLIS data, and had at least one test for COVID-19, comprising 13.2% of persons using shelters in the GTA. of these 1398 individuals who may use shelters, 6.4% tested positive, comprising 24% of all diagnosed cases of COVID-19 in shelters by May 20, 2020.**"
We have provided the rational of removing results pertaining to testing for persons using shelters in the 1ˢᵗ paragraph of the Results section on page 7:
"OLIS captured/identified 92%, 97%, and 24% of confirmed cases in all settings combined, LTCH residents, and persons using shelters, respectively, compared to iPHIS (**Appendix-2 Figure-1A-C**). **Given low sensitivity of OLIS data in identifying persons using shelters, we did not report results on testing for persons using shelters**."
We have also included the following sentence in the limitation section of the Discussion on page 10:
"Fourth, **we could not estimate testing per capita or test positivity proportion among persons using shelters given low sensitivity in identifying this population in the testing data (26). However, our data suggest a minimum of 13.2% of persons using shelters had been tested by May 20, 2020 (Appendix-2 Figure-1C), suggesting that higher diagnoses per capita among persons using shelters may be partially explained by increased testing. Work is underway to improve the sensitivity of algorithms to identify persons experiencing homelessness (76).**"
3. Thus the main conclusion appears to be that the outbreak was worse in LTC. This is already known but the additional detail in this paper would still be valuable.
We thank the reviewer for recognition of the value of the details presented in the current study.
4. When presenting that LTC cases were more likely to be female it would be helpful to see if this simply reflects that their baseline population is more likely to be female. Similarly, the higher rate of case positivity in men in shelters may reflect that their populations are more skewed to men.
Due to the lack of specific data on the population distribution by sex for LTCH residents and persons using shelters, we presented the average diagnosed cases per capita for each population/setting without adjusting for age and sex.
However, we have added the following sentence as a footnote to Table 1 where the potential distribution by sex of diagnosed cases was noted citing literature:
"**ᵈThe over-representation of females among diagnosed cases of LTCH residents should not be interpreted as increased risk of diagnoses for female residents; as data among all LTCH in Ontario has shown that more than 75% of LTCH beds are occupied by females (77). Similarly, around 43.4% of persons using shelters are female (Appendix-1); and 50.9% of the GTA population are female (3).**"
5. While the authors have done a good job of identifying limitations, another major limitation is lack of detail on co-morbidities. The worse outcome in LTC may be more due to co-morbidities than it being a congregate setting. In fact the data does sort of point to this as the case-fatality rate was lower in the shelters.
We have edited the following sentences in the 4ᵗʰ paragraph of the Discussion section on page 10:
"**Similarly, the wider scope of testing for LTCH residents could lead to larger proportion of diagnoses of infected cases. Therefore, the infection fatality rate ratio may be even higher than the 1.4 times case fatality rate ratio observed in the current study between LTCH residents versus the rest of the population.** The higher age- and sex-adjusted case fatality rate among LTCH residents as compared with the rest of the population, may reflect underlying differences in comorbidities associated with COVID-attributable mortality and/or goals of care (64). **Future studies including information on comorbidities could help identify causal pathways between residing in LTCH and increased case fatality rate.**"

6. What is interesting is the dynamic nature of the outbreak over time. Information that would be valuable would be the changing case fatality rate over time in the various settings. I haven't seen this data presented anywhere and would be interested in seeing this if it can be calculated. Obviously this will be affected by changing testing rates but it might help individuals better conceptualize how the outbreak is progressing.

We have added analysis to calculate case fatality proportions over time by setting. We have added the following sentence to describe the method on the 3rd paragraph of the Analyses subsection of the Methods section on page 5:

"Third, we calculated **the following measures over time** in the three settings for which we had data on the population size (LTCH residents, persons using shelters, and the rest of the population): cumulative diagnoses per capita, cumulative proportion of population tested, daily and cumulative proportion of individuals who tested positive, and the cumulative case fatality proportion. **For the case fatality proportion over time, a rolling average of 7 days was computed using the center method (38)."**

We have also added the following sentence to the last paragraph of the Results section on page 8, including a newly added Appendix-2 Figure-4 below:

"The age- and sex-adjusted case fatality rate was 1.4 times (95% CI: 1.1-1.8) higher among LTCH residents compared with the rest of the population (Table-1) ; **and the age- and sex-specific case fatality rate ratios ranged from 1.2-7.6 (Appendix-2 Table-3). The case fatality proportion remained relatively stable over time from April 15th onward for all settings (Appendix-2 Figure-4)."**
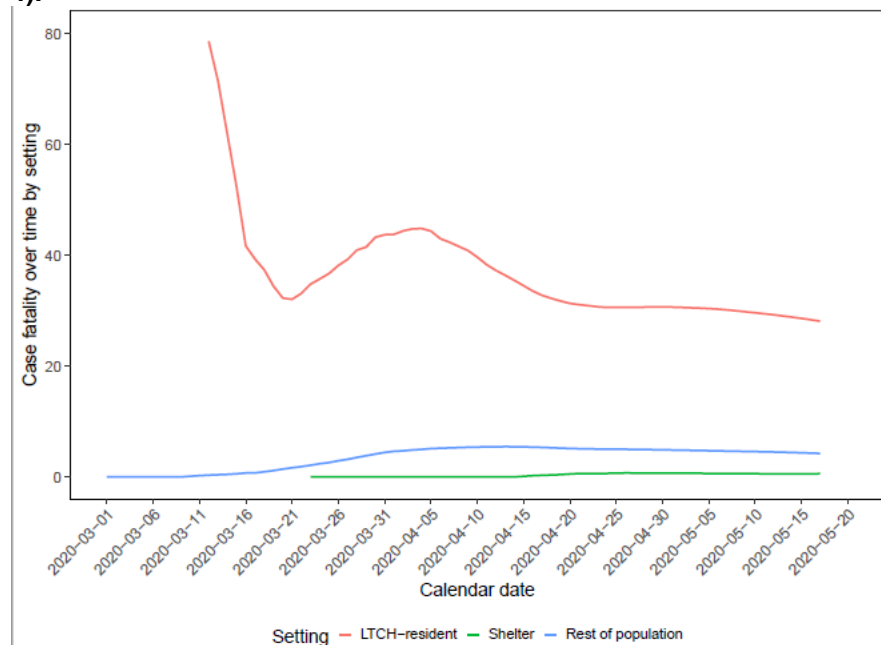
Appendix-2 Figure-4. Cumulative case fatality proportion over time in the Greater Toronto Area by outbreak setting. The calendar date refers to the date when the case was reported to the public health unit; and the follow-up was till May 31, 2020 to determine the outcome of the cases. A rolling average of 7 days was computed using the center method (e.g., to compute the case fatality proportion of any cases reported up to a certain date, the data 3 days prior and after were used to compute the average). Data sources: iPHIS: integrated Public Health Information System. Abbreviations: LTCH, long-term care homes.