

**Article details:** 2022-0022

**Title:** The association of facility ownership with COVID-19 outbreaks in long-term care homes in British Columbia, Canada: a retrospective cohort study

**Authors:** Michelle Cox, Margaret McGregor, Jeffrey Poss, Charlene Harrington

**Reviewer 1:** Vincenzo G. Menditto

General comments (author response in bold)

This retrospective study investigated the relation between facility ownership and staffing characteristics of long-term care facilities and COVID-19 outbreaks in a large region of Canada. The article is well written. I suggest a minor revision of the manuscript to publish it:

1. The study population would be better characterized, in particular about the vaccinal status of the residents. Little is also described about the use of personal protective equipment among the staff.

**Methods: Data sources, Page 3, 1st paragraph:**

**We have added the following line regarding vaccination status to the Methods section:**

**“At the beginning of the study, all residents and staff were unvaccinated. In British Columbia, COVID-19 vaccines were administered to LTC residents and staff beginning December 2020.”**

**Regarding personal protective equipment among staff, similar to other studies, there were no data available on this.**

2. In the "Limitations" section the authors wrote that definitions of outbreak can vary across studies. If possible, the authors should better specify if the laboratory-test for COVID-19 diagnosis were prescribed only in symptomatic residents/staff or also for screening.

**Methods: Data sources, page 3, 1st paragraph:**

**We have added the following sentences regarding laboratory testing for COVID-19:**

**“A PCR test was administered if a resident met the criteria for testing in accordance with provincial testing guidelines for long term care. These guidelines initially recommended testing for those with a defined set of COVID-19 symptoms, but were updated over time as public health developed a better understanding of the atypical nature of disease presentation in the frail elderly.”**

3. The article is a retrospective study, but I think that a case-control study would be more appropriate. I suggest commenting on this choice.

**While a case-control study would also be an appropriate option and is a good option for future studies, we feel our analysis is appropriate due to resource limitations. Our approach has also been used in other studies.**

4. Living facilities, supportive housing, group homes, palliative and hospice homes were excluded and it could limit the generalizability of the results: I suggest to add/stress this point in the "Limitations" section.

Limitations, page 9, 3rd paragraph:

**We have add the following sentence to address this limitation:**

**“The generalizability of the results are further restricted to specifically LTC facilities since all other types of homes, such as assisted living, were excluded.”**

**Reviewer 2:** Jesse Knight

**Institution:** Institute of Medical Science, University of Toronto

General comments (author response in bold)

The authors examine institution-level risk-factors associated with COVID-19 outbreaks in long-term care facilities in BC during epidemic waves 1-2. The motivation for the analysis is strong and the overall structure of the paper is very good. A few minor details of the methods can be clarified. The main limitation of the paper relates to variable selection, which unfortunately has the potential to invalidate some of the main findings.

#### Major Issues

1. Variable selection: Several steps of the variable selection process are potentially problematic. First, pre-screening variables by bivariate analysis, although popular, is an invalid approach that risks rejecting important variables that are confounded before multivariate analysis (see: Sun, Shook, & Kay (1996) JCE). Then, it seems a similar step is used with "category models", and thus may similarly bias effect estimates. Incidentally, no mention of category models is made in the results, and their interpretation versus fully adjusted model is not discussed, so likely they should be omitted altogether. Finally, selection of only the "most significant" variables from each category to use in the final model is not justified either. Such variables may be related but independent predictors of the outcome. Addendum: Table 3 \* footnote describes another similar issue with removing case mix index. Heinze, Wallisch, & Dunkler (2018) *Biometr J*, provide helpful guidance on variable selection problems, (esp. Table 3). Given the number of variables is  $\sim 1/10$  the number of outbreaks ("events per variable" EPV  $\sim 10$ , depending how other issues raised below are resolved), either of the following could be justifiable: a) no variable selection; b) variable selection among some variables (those less certain), followed by shrinkage and stability analysis as described by Heinze. Perhaps some variables could also be aggregated/combined. e.g. total direct, nursing, and allied care hours per resident day could perhaps be summed to yield one variable; the three sub-contracted services could counted as the (ordinal) number of sub-contracted services (0,1,2,3); and one of "any shared beds" and "large % shared beds" could be selected a priori. Note that bivariate analyses could still be included, but should not be used for variable selection.

**We have revised the way we selected our variables for our multivariable analysis. We have now selected our variables for the multivariable model a priori based on the current literature describing the association of staffing levels, facility size and age, and the presence of shared rooms on outbreaks. This is now described in the methods section.**

Methods, Statistical analysis, page 4, 3rd paragraph:

**"In addition to ownership, explanatory variables for the Cox proportional hazards analyses were determined a priori from a review of the literature and in consultation with clinical experts in LTC. If variables were highly correlated then only one variable of interest was included."**

#### Moderate Issues

2. Outbreak definition includes staff cases: By definition, facility outbreaks must import infection from the community, via either staff or visitors. Therefore, a single staff case does not indicate transmission within the facility. Even multiple concurrent staff cases could be simultaneously acquired in the community. Therefore, the main exposures of interest -- facility-level characteristics -- would not be expected to explain most "outbreaks" (by this definition) comprising only staff cases. Arguably, the important risk to quantify is risk of transmission to residents alone, who are vulnerable to worse outcomes, and whose cases could be better predicted by the exposures of interest. Refs Li [10], Stall [13], He [19] all examine only resident cases, while Gorges [24] included both residents and staff, and Abrams [12] was not clear. As such, the impact of the paper could be improved by considering only resident cases. This may also avoid bias towards the null from attempting to explain staff cases that are not driven by the variables explored. If including both residents and staff is well justified, then perhaps a sensitivity analysis could explore resident-only outbreaks.

**We have revised our outbreak definition (described in previous responses) to include resident cases only and have excluded any staff cases. Tables 2 and 3 are now reflective of this revised definition.**

3. The current analysis ignores all information about the number of cases associated with each outbreak for quantifying risk from each exposure. This number is perhaps a better indicator of facility-level effectiveness at preventing internal transmission, versus any 1+ case, even among residents, which may be acquired from visitors. The hurdle model approach by Stall et al. [13], which separately quantifies risk of any single case or "outbreak" (logistic), and then total number of cases (quasi-poisson) given an outbreak is a nice way to consider both aspects. This would be a great addition to the paper, as it appears the required data are available; then, risk factors related to infection entry vs within-facility transmission could be distinguished and better understood.

**Due to the limited funding and resources for this study, we were unable to look at outbreak extent in addition outbreaks.**

4. The current description of variables in the methods section presents levels of a categorical variable (ownership) using the same notation as separate variables that are grouped together (staffing, sub-contracted), which made it difficult to easily identify unique variables: "Explanatory variables interest included facility ownership (for-profit, non-profit, or health authority owned and operated), staffing levels (total direct, nursing, and allied care hours per resident day), and sub-contracted out services (professional and non-regulated nursing, and food services)." I had to read Table 2 to understand the variables used. Please consider rephrasing for clarity.

Methods: Outcome and explanatory variables, page 3, 2nd paragraph:

**We have reworded this sentence to clarify the unique variables, it now reads: "Explanatory variables of interest included: facility ownership (for-profit, non-profit, or health authority owned and operated); two continuous staffing level variables (total nursing/care aide hours per resident day and allied health hours per resident day); and three separate dichotomous sub-contracted out services variables (professional, non-regulated nursing, and food services)."**

5. The tests for significance in bivariate analyses are not described in the methods, only listed in Table 2 footnote. These tests should be noted in the text, and more clearly indicated which tests were applied to which variables.

Main manuscript, page 4, 2nd paragraph:

**We have added the statistical tests used for the bivariate analysis to the statistical methods section. It reads as follows:**

**“Tests of comparison for the bivariate analyses included Welch’s t-Test or Wilcoxon-Mann-Whitney test for continuous variables, and Chi-square test for categorical variables.”**

Minor Issues

6. The phrase "repeated events" or similar could be added to the methods help clarify this specific extension of CPH. Currently, the use of this extended CPH model is only alluded via "Facilities were deemed at risk of an outbreak throughout the duration of the study ... except when experiencing a declared outbreak". and later "correlation among observations between ... subsequent outbreaks at the same facility".

Methods, Statistical Analysis, page 4, 3rd paragraph:

**We have included an additional sentence to clarify this specific extension of CPH:**

**“Facilities could experience more than one outbreak, and contributed a new observation starting the day after an outbreak was over.”**

**Methods, Statistical Analysis, page 5, 1st paragraph:**

**We have also replaced the word ‘subsequent’ with ‘repeated’ to further clarify. So the sentence now reads:**

**“.....and included a robust sandwich estimator for the covariance matrix to account for correlation among observations between health authorities and repeated outbreaks at the same facility.”**

7. The details of the sensitivity analysis should be added to the methods, namely how outliers were identified and what threshold for "proportionally fewer outbreaks" was used to exclude some health authorities.

Methods, Statistical analysis, page 5, 2nd paragraph:

**We have added details of the sensitivity analyses to the methods section. The paragraph now reads:**

**“Sensitivity analyses were conducted to test the robustness of the final model results, by re-running several different models. The alternate models excluded either outlier facilities or health authorities with fewer outbreaks, or used a more restrictive outbreak definition (did not include outbreaks with only a single resident case). Outlier facilities excluded for the sensitivity analyses were identified by visual inspection and by graphing boxplots. Health authorities with fewer outbreaks were identified by visually inspecting the data.”**

8. missing word? "a rolling 14-day (average?) incidence"

Methods, Outcome and explanatory variables, page 4, 1st paragraph:

**Thank you, we have added this word.**

9. Table 2 "ownership characteristics" includes several re-stratifications of the variable that are not described in the text, such as "contract" vs authority, and chain/multi-site characteristics. These should either be described in the methods or removed from Table 2.

Table 2:

**Thank you, we have removed these additional stratifications.**

10. Table 2: the symbols § and † appear to be mixed up in the final 3 rows.

Table 2:

**Thank you, we have corrected this error.**

11. grammar: "there (was -> were) 164 COVID-19 outbreaks"

Results, COVID-19 outbreaks, cases, and deaths, page 5, 1st paragraph:

**Thank you, we have corrected this grammatical error.**

12. "The mean total direct care" Why are only direct care and nursing hours reported, not allied health?

Table 2 and Results, Long-term care facility characteristics, page 6, 2nd paragraph:

**Allied health hours are reported in Table 2 under Staffing Characteristics. We have also added this result to the Results section which now reads:**

**"The mean total nursing/care aide, and total allied health hours across all facilities were 3.0 and 0.3 hours per resident day, respectively."**

13. Consider adding "Bivariate analysis:" to the heading "Characteristics by outbreak status" and likewise "Multivariate analysis: Associations with risk of outbreak" as signposting for the reader

Table 2 & 3:

**Thank you for the suggestion, we have added the Bivariate analysis and Multivariable analysis to Tables 2 and 3, respectively.**

14. The sentences "Comparisons of facilities with an outbreak versus those with no outbreak over the study period are presented. There were significant differences in ownership characteristics between the two groups." can be removed as they are somewhat redundant with the subsequent specific results. Perhaps instead, the p-values indicating significance can be added to each reported bivariate result in the main text.

Results, Characteristics by outbreak status, page 6, 3rd and 4th paragraphs:

**We have removed these two sentences and added the p-values to the reported bivariate results in the text.**

15. Has incidence for the CPH model been transformed to per-1000? The additional hazard (HR-1) is very small and difficult to interpret. Consider re-scaling the variable to improve interpretability.

Tables 2 & 3:

**We have now transformed the incidence for the CPH model to per 100,000 to improve the interpretability.**

16. The results sentence on sensitivity analysis refers to Table 3, but no results of sensitivity analysis are presented there?

Results, Association with risk of outbreak, page 7, 2nd paragraph:

**We have moved the referral to Table 3 and added "data not shown" to clarify which results are presented. The sentences in this paragraph now read:**

**"Facilities with shared rooms were not at higher outbreak risk compared to those with no shared rooms (Table 3). For the sensitivity analyses with facility outliers removed, data restricted to health authorities with higher community incidence of COVID-19, and models using a more restrictive definition of outbreak (more than one resident case) the direction of effect estimates remained unchanged (data not shown)."**

17. "both of which are key (strategies -> resources)" as PPE are not really a "strategy"

Interpretation, Explanation of findings, page 8, 2nd paragraph:

**We have made this change. The sentence now reads:**

**".....and personal protective equipment (2), both of which are key resources for outbreak prevention."**

18. "this lost significance in the multi-variable model" -> "this effect was not significant after adjustment"

Interpretation, Explanation of findings, page 9, 2nd paragraph:

**Thank you, we have actually removed this sentence. The sentence now reads:**

**"Our study found that higher hours of nursing/care aide hours per resident day were protective in the univariate model, although this effect lost significance in the adjusted model."**

19. "associated with (odds -> risk) of an outbreak" as the model used estimated hazards, not odds.

Interpretation, Explanation of findings, page 9, 2nd paragraph:

**Thank you, we have made this change.**

20. How the proposed directions of future work related & build upon the current findings could be made more clear.

**We addressed this comment but adding to the section regarding future directions.**

**Interpretation, Future directions, page 9, 2nd paragraph now reads:**

**"Future research to build on the current findings should deploy more qualitative methods to better understand the multiple underlying pathways that may have contributed to the protective effect of public ownership. Interviews with frontline staff, residents and facility directors of care could explore factors including consistency of personal protective equipment practices, the proportion of facility funding allocated to staffing, timely implementation of single site orders, infection control and prevention policies, team cohesion, other measures of facility leadership such as staff trust, and directors of care length of employment as potential contributors to outbreak risk. Understanding the association of these factors with both outbreak risk and with facility ownership will help inform meso and macro-level LTC policy and better prepare facilities for future similar events. Further research from BC and other provinces is also needed to examine the association of facility ownership and other characteristics with outbreak severity and mortality."**

21. "health authority ownership ... was protective ... the factors behind this are likely multidimensional" This phrasing suggests that the other variables explored could be mediators of the main effects. However, no mediation analysis has been done -- in fact the opposite as other variables were adjusted. Consider rephrasing.

**We have now rephrased this paragraph.**

**Conclusion, page 10, 2nd paragraph now reads:**

**"The study findings suggest that health authority ownership of LTC facilities in British Columbia was protective of COVID-19 outbreaks. Further study is needed to unpack the underlying pathway(s) behind the observed association to inform policy for mitigating the negative impact of future outbreaks on this vulnerable population."**

22. The limitations section concludes, "our study is the first Canadian study to examine outbreak risk by ownership, staffing, and a number of other facility characteristics." The 2020 study by Stall [13] examined many of the same factors, so perhaps this can be rephrased.

Limitations, page 10, 1st paragraph:

**This line now reads:**

**“Despite these limitations, our study adds to the limited Canadian research on this topic in a context that is arguably very different from Ontario when it comes to LTC facility ownership.”**

23. Consider replacing all instances of "utilize" with "use" to improve readability.

**Thank you for your suggestion, these have all been replaced.**

24. Citations and Table call-outs are inconsistently before/after punctuation. They should all appear before the punctuation.

**Thank you, we have made this change so all call-outs are now consistently before the punctuation.**

**Reviewer 3:** Soo Chan Carusone

**Institution:** Casey House

General comments (author response in bold)

The paper “The association of facility ownership and staffing characteristics with COVID-19 outbreaks in long-term care homes in British Columbia, Canada: A retrospective cohort study” is well timed and adds important data and analysis to necessary conversations. Overall, I think the paper is well written. I have some questions and comments, most specifically related to the methods and reporting of the methods. I have organized my comments below by location in the paper and then by importance, with some more minor considerations at the end.

Major comments:

1. On page 4, you state that a “decision to end the study on January 31, 2021 was made since by that time virtually all residents and a higher number of long-term care staff had been fully vaccinated” – did this correspond to a significant decrease in outbreaks? Presenting the trend in outbreaks may be more appropriate/relevant; or if it was done for administrative or other reasons, I suggest removing this sentence.

Methods, Data sources, page 3, 1st paragraph:

**We have added to this sentence to clarify why the study was ended. The sentence now reads:**

**“A decision was made to end the study on January 31, 2021 since both the number and severity of COVID-19 infections substantially decreased following full vaccination of the LTC population around this time.”**

2. In the outcome and explanatory variables and statistical analysis section, please state whether these variables and the analysis plan was determined a priori and also how the variables were defined and analyzed (e.g., dichotomous vs categorical). I am also curious, it seems like you dichotomized sub-contracts for services; are there circumstances where facilities employ their own staff but supplement with sub-contracted staff (e.g., mixed organization employed and agency nurses)?

Methods, Outcome and explanatory variables, page 3, 2nd paragraph:  
**We have indicated which variables are dichotomous and categorical.**

Methods, Statistical analysis, page 4, 1st paragraph:  
**We have added a sentence to indicate how the variables were analyzed, it reads as follows:**

**“Tests of comparison for the bivariate analyses included Welch’s t-Test or Wilcoxon-Mann-Whitney test for continuous variables, and Chi-square test for dichotomous or categorical variables.”**

**We agree there are circumstances that exist where facilities employ their own staff but supplement with sub-contracted staff. However, in this case we are referring to sub-contracting of an entire work force job classification, rather than the temporary use of agency nurses to fill a shift that is short-staffed.**

3. On page 5, please state the significance level threshold used for inclusion in the multi-variable analysis.

Statistical analysis, page 5, 1st paragraph:

**As suggested by the reviewers, we have revised the way we selected variables for our multivariable model. The sentence in the statistical analysis section now reads:**

**“In addition to ownership, explanatory variables for the Cox proportional hazards analyses were determined a priori from a review of the literature and in consultation with clinical experts in LTC. If variables were highly correlated then only one variable of interest was included.”**

4. Is there a methodological reference you could provide for your choice to only include the most significant variable in a category in the multi-variable analysis (when more than one was significant in bivariate analyses). I also wonder if a discussion of the significance of this on the limitations or interpretation of the findings should be included later in the paper.

Statistical analysis, page 5, 1st paragraph:

**We have revised the way we selected variables for the multivariable analysis and have described this in previous responses above.**

5. In the summary of results section I suggest rewording the findings to indicate where you are referring to results from multi-variable regression models (and what the other variables were).

**We have added multivariable and bivariate to the respective paragraphs in the results and main findings sections to indicate which results we are referring to.**

Interpretation, Main findings, page 7, 1st paragraph:

**We have added multivariable analysis to this paragraph to indicate that we are referring to the multivariable regression models.**

**“In the multivariable analysis, we included 293 LTC homes and found that ownership by a health authority was protective of a COVID-19 outbreak compared to both for-profit and non-profit facilities.”**

Results, Characteristics by outbreak, page 6, 1st paragraph:

**The paragraph begins as follows:**

**“In the bivariate analysis, among those facilities with an outbreak, 46% were for-profit, 31% were non-profit and 23% were health authority owned and operated.”**



Results, Associations with risk of outbreak, page 7, 2nd paragraph:

**The paragraph now begins as follows:**

**“In the multivariable regression model, for-profit and non-profit ownership were associated with a higher risk of COVID-19 outbreak compared to health authority owned and operated facilities, adjusted HR 1.988 (95% CI 1.122, 3.523) and 1.837 (95% CI 1.004, 3.363) respectively. The model adjusted for regional incidence of COVID-19 (adjusted HR 1.118 (95% CI 1.072, 1.166) and higher number of beds in a facility (adjusted HR 1.204 (95% CI 1.119, 1.295) per 25 beds). Higher total nursing/care aide hours per resident day was also included in the model and was inversely but non-significantly associated with risk of outbreak (adjusted HR 0.842 (95% CI 0.332, 2.135).”**

6. Table 1: Is the median resident cases per facility presented for all facilities or only the 112 that had one or more case? (The latter may be more interesting.) I also suggest putting the 25th and 75th percentile vs the IQR given the median is 8 and the IQR is 40.

Table 1:

**We have now presented the mean resident cases per facility with the standard deviation. It is presented for only those facilities that had one or more resident cases (an outbreak), N=80.**

7. I am wondering if the variables impacting staff only outbreaks (especially where there was only one case), may be quite different, and thus not appropriately pooled with resident involved outbreaks (e.g., SES of surrounding neighbourhood/staff). I realize there is significantly less power if you exclude these but did you do any sensitivity analyses looking at this? What has been done in other studies?

**We have revised our definition of outbreak to include only resident cases and reanalyzed the data for our bivariate and multivariable analyses as described in previous responses.**

8. In BC, are there other known factors that may differ between ownership-type? E.g., urban vs rural, SES of neighbourhood, affiliation/co-location with another health care facility?

**Thank you for your comment. For this study, most outbreaks were in sub-urban and urban areas. There were very few outbreaks in rural areas; therefore, we did not include a variable that measured this. We did measure community prevalence, and SES and community prevalence are highly correlated.**

9. I agree with many of your areas for future research but suggest you provide some framing for the reader, are you suggesting looking at how these are individually associated with outbreak risk and/or how these may differ by ownership-type?

**We are suggesting both – the former might help to inform LTC policy in individual care homes and/ or improved contact management and possibly funding by health authorities for contracted services delivery. The second might suggest a health policy that favours health authorities assuming a greater share of ownership of long-term care facilities. We have expanded this section to improve the framing of future directions.**

Additional comments:

10. Under data sources, I believe it should be “Canadian Institute for Health Information (CIHI)”

Data Sources:

**Thank you, we have corrected this.**

11. The journal may have guidelines for the number of decimal places to report but I would suggest that in many cases within the text fewer would be more appropriate given the data and interpretation e.g., percentage of homes by ownership and two decimal places in the number of hours of direct care. I don't think the accuracy of reporting is sufficiently robust to make interpretation of differences of 0.01 hours/resident bed day appropriate.

**We have removed the decimal places for all percentages reported in the text. We have only included one decimal places for all hours per resident day reported in the text.**

12. I am not a statistician but wonder if you should acknowledge the interactions between the variables included – perhaps in the limitations and/or next steps section.

**We have added a sentence to the limitations section to acknowledge the interactions between variables included.**

Interpretation, Limitations, page 10, 1st paragraph:

**"We acknowledge that hazard ratios of multivariable models express an average effect across the range of covariate values, and that interactions not explored here could show other findings specific to certain measure combinations."**

13. Table 1: I assume the deaths counted are only residents, but it may be helpful to specify this.

Table 1:

**We have added a row to the Table 1 to clarify the number of resident deaths. There were no staff deaths due to COVID-19 in long-term care in this province.**

14. Table 1: I find the heading “mean days of outbreak per outbreak” rather confusing. Would it be simpler to say “mean length of outbreak”?

Table 1:

**We have made this change.**