PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cohort profile: the British Columbia COVID-19 population mixing patterns survey (BC-Mix)
AUTHORS	Adu, Prince; Binka, Mawuena; Mahmood, Bushra; Jeong, Dahn; Buller-Taylor, Terri; Damascene, Makuza Jean; Iyaniwura, Sarafa; Ringa, Notice; Velásquez García, Héctor A.; Wong, Stanley; Yu, Amanda; Bartlett, Sofia; Wilton, James; Irvine, Mike A.; Otterstatter, Michael; Janjua, Naveed

VERSION 1 – REVIEW

REVIEWER	Coletti, Pietro Universiteit Hasselt, Censtat
REVIEW RETURNED	09-Nov-2021

GENERAL COMMENTS	In the manuscript "Quantifying contact patterns: development and characteristics of the British Columbia COVID-19 population mixing patterns survey (BC-Mix)" the authors introduce the BC-Mix survey, presenting its study objective, the methodology used and briefly discuss the collected sample size. They refer to future publications for the survey results.
	The survey protocol is, on average, well described and only minor adjustments are needed for publication. I do not have major comments and I list below some minor comments that could improve the manuscript readability and would include some relevant literature that could be useful for the reader. All in all, I think that the BC-Mix has been properly designed and that it will provide interesting results, considering also its large sample size.
	Some minor comments (the page number used is the one showing in the upper-right corner of the review file, listing up to 27 pages): - Page 5, line 52 " Early detection of COVID-19 resurgences requires mechanisms for tracking precursors of transmission, including changes in social contacts, mixing patterns and physical distancing behaviours as well as early signals of a COVID-19 spread." As it is written the sentence seems to imply that changes in social contacts can be used as precursors of transmission. Although this could be in principle true (individuals are known to change their behavior when experiencing symptoms) the effect is probably too small to be used as a precursor of transmission. Indeed collection of social contact data can help in navigating the pandemic, but for its predictive value on the spreading potential. I would amend the sentence to remove the link between social contacts and monitoring precursors of transmission. - Page 6, line 10 "These population-specific data could [...]rather than inferring these from reported cases and hospitalizations (6,6)": Amend the double citation and consider also including other examples, like Zhang et al., Coletti et al. , or others.

- Page 6, line 15 "Various studies have assessed the impact of physical distancing measures[...]": I think that the contribution from Liu et al. (a rapid review on the topic) and from Verelst et al. (presenting the European CoMix study) should be acknowledged.

- Page 6, line 49 "To capture participants from a broad demographic range, the survey invitation and survey are disseminated through Instagram [...]": I would remove <and survey>, as it is probably the survey invitation that is disseminated.

- Page 6, line 55 "We also monitor the demographic profile of survey participants and occasionally use these functions to target recruitment to age groups or sex that may be under-represented (16).": Could the author be more specific about this "occasional use"? Is there a defined protocol? Is it based on a scientific assessment of the under-representation or are there defined some threshold values?

Page 8, line 26 "The widely-used HBM, has previously been used to evaluate beliefs and attitudes toward seasonal influenza and pandemic swine flu vaccines as well as the COVID-19 vaccine (27–29).": The authors could consider including the recent work on Wambua et al. on the HBM in relation to vaccination and number of contacts.

Page 8, line 55 "All duplicates are removed": could the authors present an example of duplicate removed? Do they mean duplicate responses from the same participant (i.e. filling twice the survey in the same time window)?

Page 11, line 11 "Quota sampling has been used by other studies to achieve representativeness (7,44). We used two approaches to achieve the same goal: adaptive recruitment through promotion and targeting to specific populations and then post hoc weighting. Our survey tool does not set quotas on recruitment but uses targeted advertisements to improve representativeness. ": I would remove this sentence from here and insert this in the "Analysis, data cleaning and weighting" section.

Page 11, line 25 "Also, people who are in prison (sentenced or on remand) or people who are under immigration detention may not have access to the internet or cellular devices.": This is indeed a limitation of the study, but why is it relevant for a population-based study? I would think that the number of individuals for which this applies is not massive. Do the authors intend to specifically address the prison/immigrant population? Why, and how, are these populations relevant for COVID-19 transmission in the population?

- Page 12, line 55 "[...] many important variables besides age and sex [...]": Would these be geography and ethnicity? In this case I would list them explicitly, instead of referring to them as <many>. If this is not the case, then also update the methodology section.

References

Coletti et al. "A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies". BMC Infectious Diseases (doi: https://doi.org/10.1186/s12879-021-06092-w).

Liu et al. "Rapid Review of Social Contact Patterns During the COVID-19 Pandemic". Epidemiology (doi: 10.1097/EDE.0000000000001412).
Verelst et al. "SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries". BMC Medicine (doi:https://doi.org/10.1186/s12916-021-02133-y).
Wambua et al."The influence of risk perceptions on close contact frequency during the SARS-CoV-2 pandemic". Research square (doi: https://doi.org/10.21203/rs.3.rs-996488/v1).
Zhang et al. "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China". Science (doi : 10.1126/science.abb8001).

REVIEWER	van der Sande, Marianne A. B.
	University Medical Centre Utrecht
REVIEW RETURNED	04-Feb-2022

GENERAL COMMENTS	This study describes the set-up of an open, ongoing cohort to collect data on (changes in) contact patterns. This is not a novel concept and builds on similar studies before and during the Covid-19 pandemic. Results are not yet presented, but a great number of analysis in relation to specific Covid-19 interventions is announced.
	To understand human-to-human transmission dynamics of infectious diseases, including assessment of interventions aiming to impact on such transmission, it is highly valuable to have real time representative data on (changes in) contact patterns. However, there remain several concerns and unclarities which need to be addressed:
	-Records were accepted if at least 33% complete: please add justification for this cut off, and discuss implications for such a low threshold for validity of surveys. Also, no data are given on % complete by survey round, or by population group, please add and discuss. Currently it is unclear how complete entries were, if incompleteness was random with respect to questions, population groups, survey rounds/time of follow-up, etc. Please add at least a sensitivity analysis with a cut-off of at minimum 67%?
	-Explain why sex, but not geography and ethnicity were required to be included. If geography was not available, how have the indicators at area level been generated and validated? Also, is it correct that for sex the only options were male and female, and that people who do not define as such were excluded?
	-Table 1 shows large discrepancies between cured and weighted frequencies, suggesting rather biased uptake of the survey. Also, the unweighted sample clearly differs from the BC population it should represent. Please explain much more clearly the rationale and methodology of the weighing, and discuss implications for validity and representativeness of the large differences. Eg only 16.5% of participants were male, which has been artificially adjusted to a 50% weighting (even beyond the male population share of 48.5%). Does this mean that the data from the highly selected group of male participants are considered as informative as the slightly less selected group of women? The statement that weighing and using many auxiliary variables to fill considerable selection gaps increased the representativeness is not supported by data, and could be wishful thinking.

-The study started in September 2020 and data are reported up to August 2021. This would suggest that participants who joined in the first month would have at least 11 follow-up measurements, but no more than 8 follow up surveys are reported (with only 4% of the baseline sample participating in the 8th round).

It is unclear to what extent this is related to attrition, data management problems, survey design factors, ..but adds to concerns to what extent valid longitudinal (monthly) data are available. Also, it it is not clear to what extent people in different rounds were the same people (as it should be in a cohort) or more like repeated X-sectional surveys among partially overlapping subgroups.

Please clarify in full the recruitment (how many people started per month?) and follow up process of each month of the cohort; also please discuss implications of any differences in characteristics between people in different recruitment trajectories.

-It is unclear why only approx.. 2/3 of survey responses not only in the baseline round but also in each follow-up round were considered eligible (please add % for this throughout to facilitate), suggesting some serious design issues.

-It is unclear how the change in number of contacts included (first three, then ten) affects interpretation of data, and how this is/will be dealt with

All of the above points suggest multiple sources of selection bias, at recruitment and (increasing) over time,
and raises significant concerns on the validity of any responses received and imputed.

The conclusion that this tool provides critical data for timing of interventions is unsubstantiated, and without proper validation of representativeness of contact data rather doubtful in view of high attrition, low completeness, significant selection bias.

Some minor comments:

-in the questionnaire, please clarify what variables were asked only at baseline (with an update after x months or not?), what variables (also) in the follow-up rounds

-several sloppy mistakes: eg line page 4 line 39 states that the survey began September 2021 (rather than 2020)

VERSION 1 – AUTHOR RESPONSE

<u>Reviewer: 1</u>
Comments to the Author:
In the manuscript "Quantifying contact patterns: development and characteristics of the British Columbia COVID-19 population mixing patterns survey (BC-Mix)" the authors introduce the BC-Mix survey, presenting its study objective, the methodology used and briefly discuss the collected sample size. They refer to future publications for the survey results.
The survey protocol is, on average, well described and only minor adjustments are needed for publication. I do not have major comments and I list below some minor comments that could improve the manuscript readability and would include some relevant literature that could be useful for the

reader. All in all, I think that the BC-Mix has been properly designed and that it will provide interesting results, considering also its large sample size.

RESPONSE: Thank you for the compliments.

Some minor comments (the page number used is the one showing in the upper-right corner of the review file, listing up to 27 pages):
- Page 5, line 52 " Early detection of COVID-19 resurgences requires mechanisms for tracking precursors of transmission, including changes in social contacts, mixing patterns and physical distancing behaviours as well as early signals of a COVID-19 spread." As it is written the sentence seems to imply that changes in social contacts can be used as precursors of transmission. Although this could be in principle true (individuals are known to change their behavior when experiencing symptoms) the effect is probably too small to be used as a precursor of transmission. Indeed collection of social contact data can help in navigating the pandemic, but for its predictive value on the spreading potential. I would amend the sentence to remove the link between social contacts and monitoring precursors of transmission.

RESPONSE: We agree with this comment. We have removed this sentence (page 3, paragraph 3).

- Page 6, line 10 "These population-specific data could [...]rather than inferring these from reported cases and hospitalizations (6,6)": Amend the double citation and consider also including other examples, like Zhang et al., Coletti et al. , or others.

RESPONSE: Thank you for the suggestion. We have corrected the citation and have included the suggested citations (page 4).

- Page 6, line 15 "Various studies have assessed the impact of physical distancing measures[...]": I think that the contribution from Liu et al. (a rapid review on the topic) and from Verelst et al. (presenting the European CoMix study) should be acknowledged.

RESPONSE: Thank you for the suggestion. We have acknowledged the suggested studies. (Page 4)

- Page 6, line 49 "To capture participants from a broad demographic range, the survey invitation and survey are disseminated through Instagram [...]": I would remove <and survey>, as it is probably the survey invitation that is disseminated.

RESPONSE: Thank you for the suggestion. We have amended this sentence accordingly. (Page 4).

- Page 6, line 55 "We also monitor the demographic profile of survey participants and occasionally use these functions to target recruitment to age groups or sex that may be under-represented (16).": Could the author be more specific about this "occasional use"? Is there a defined protocol? Is it based on a scientific assessment of the under-representation or are there defined some threshold values?
RESPONSE: Thank you for the comment and suggestion. Our targeted recruitment strategy was to make the population participating in the survey representative of the general population of BC with respect to demographics. We tweaked our recruitment target demographics based on departures from population distribution. For example, strategies included recruiting from Instagram to increase participation of younger people. We have now clarified in the "participant recruitment section" in Page 4 that in these instances we use the BC population distribution as a point of reference.

Page 8, line 26 "The widely-used HBM, has previously been used to evaluate beliefs and attitudes toward seasonal influenza and pandemic swine flu vaccines as well as the COVID-19 vaccine (27–29).": The authors could consider including the recent work on Wambua et al. on the HBM in relation to vaccination and number of contacts.
RESPONSE: Thank you for the suggestion. We have included the suggested work (Page 6).

5

Page 8, line 55 "All duplicates are removed": could the authors present an example of duplicate removed? Do they mean duplicate responses from the same participant (i.e. filling twice the survey in the same time window)?
RESPONSE: We have amended this sentence, clarifying that duplicates refer to the same participant filling the survey more than once in a survey round (Page 7).

Page 11, line 11 "Quota sampling has been used by other studies to achieve representativeness (7,44). We used two approaches to achieve the same goal: adaptive recruitment through promotion and targeting to specific populations and then post hoc weighting. Our survey tool does not set quotas on recruitment but uses targeted advertisements to improve representativeness. ": I would remove this sentence from here and insert this in the "Analysis, data cleaning and weighting" section.
RESPONSE: Thank you for the suggestion. We have moved these sentences to the "Analysis, data cleaning and weighting" section.

Page 11, line 25 "Also, people who are in prison (sentenced or on remand) or people who are under immigration detention may not have access to the internet or cellular devices.": This is indeed a limitation of the study, but why is it relevant for a population-based study? I would think that the number of individuals for which this applies is not massive. Do the authors intend to specifically address the prison/immigrant population? Why, and how, are these populations relevant for COVID-19 transmission in the population?
RESPONSE: Thank you for the questions. COVID-19 transmission has been more prevalent among people in prison at all stages of the pandemic, and within Canada and the US, movement in and out of the prison system is near constant, with majority of people in custody at any time being on remand or on short sentences. So, they are a dynamic population, with increased COVID-19 risk, but who are largely part of the 'digital divide' (being people who do not have access to a smart phone or computer and lack internet connectivity to be able to participate in our survey). Each year in British Columbia, over 30,000 people move in and out of the correctional system, which compared to overall population of 5 million is small but still an important population group. We are attempting to address this limitation through more focussed studies that are recruiting in correctional centres, however results are not yet available.

- Page 12, line 55 "[...] many important variables besides age and sex [...]": Would these be geography and ethnicity? In this case I would list them explicitly, instead of referring to them as <many>. If this is not the case, then also update the methodology section.
RESPONSE: We have amended this sentence accordingly, noting "ethnicity" and "geography" as the other variables that were considered.

References
Coletti et al. "A data-driven metapopulation model for the Belgian COVID-19 epidemic: assessing the impact of lockdown and exit strategies". BMC Infectious Diseases
(doi: https://doi.org/10.1186/s12879-021-06092-w).
Liu et al. "Rapid Review of Social Contact Patterns During the COVID-19 Pandemic". Epidemiology (doi: 10.1097/EDE.0000000000001412).
Verelst et al. "SOCRATES-CoMix: a platform for timely and open-source contact mixing data during and in between COVID-19 surges and interventions in over 20 European countries". BMC Medicine (doi:https://doi.org/10.1186/s12916-021-02133-y).
Wambua et al."The influence of risk perceptions on close contact frequency during the SARS-CoV-2 pandemic". Research square (doi: https://doi.org/10.21203/rs.3.rs-996488/v1

VERSION 2 – REVIEW

REVIEWER	Coletti, Pietro Universiteit Hasselt, Censtat
REVIEW RETURNED	06-May-2022

GENERAL COMMENTS	The authors have addressed all of my comments. I think the manuscript is ready for publication.

REVIEWER	van der Sande, Marianne A. B. University Medical Centre Utrecht
REVIEW RETURNED	26-Apr-2022

GENERAL COMMENTS	Thank you for adding clarifications on the methodology of the survey. I still have two comments. 1. Clearly, the team has made several attempts to recruit a demographically representative sample, and then used weighing techniques to correct for the remaining demographic unbalances. However, it is not correct to present the artificially constructed weighted population as a finding: 83.5% of respondents were female, not approx 50%; 58.7% of participants were 55+, not 39%; and 45.9% of them had a university degree, not 50%. This must be corrected, in particular in the abstract. 2. Furthermore, weighting to reflect the underlying demography better does not necessarily mean the weighted data on contact patterns are representative. This depends to what extent contact patterns are also determined by other unmeasured and unweighted factors. It would be good if the authors could still include this limiation and reflect on if. They should be careful not to claim that the contact patterns obtained from the sample weighted for demographic characteristics automatically provides a representative reflection of contact patterns in the underlying population.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Dr. Marianne A. B. van der Sande, University Medical Centre Utrecht
Comments to the Author:
Thank you for adding clarifications on the methodology of the survey.
I still have two comments.
1. Clearly, the team has made several attempts to recruit a demographically representative sample, and then used weighing techniques to correct for the remaining demographic unbalances. However, it is not correct to present the artificially constructed weighted population as a finding: 83.5% of respondents were female, not approx 50%; 58.7% of participants were 55+, not 39%; and 45.9% of them had a university degree, not 50%. This must be corrected, in particular in the abstract.

RESPONSE: We have revised the abstract to distinguish between weighted and unweighted proportions.

2. Furthermore, weighting to reflect the underlying demography better does not necessarily mean the weighted data on contact patterns are representative. This depends to what extent contact patterns are also determined by other unmeasured and unweighted factors. It would be good if the authors could still include this limiation and reflect on if. They should be careful not to claim that the contact patterns obtained from the sample weighted for demographic characteristics automatically provides a representative reflection of contact patterns in the underlying population.
RESPONSE: Thank you for this suggestion. We have now included this limitation.

Reviewer: 1
Dr. Pietro Coletti, Universiteit Hasselt
Comments to the Author:
The authors have addressed all of my comments. I think the manuscript is ready for publication.
RESPONSE: Thank you.