Title: Efficacy of "stay-at-home" policy on COVID-19 transmission in Toronto, Canada: a mathematical modeling study
Authors: Pei Yuan PhD, Juan Li PhD, Elena Aruffo PhD, Evgenia Gatov MPH, Qi Li MSc, Tingting Zheng PhD, Nick Ogden PhD, Beate Sander PhD, Jane Heffernan PhD, Sarah Collier MPH, Yi Tan MSc, Jun Li PhD, Julien Arino PhD, Jacques Bélair PhD, James Watmough PhD, Jude Dzevela Kong PhD, Iain Moyles PhD, Huaiping Zhu PhD

Reviewer 1: John Glasser
Institution: CCID, CDC
General comments (author response in bold)

C1. The authors use a model that distinguishes household and community transmission to investigate the impact of various measures, actual past and hypothetical future, to mitigate the COVID-19 pandemic in Toronto. The work seems sound and results are interesting, particularly the comparison of model- and report-based reproduction numbers and thresholds for safely reopening the economy.
But the manuscript should have been edited by a co-author familiar with medical journals whose native language is English. Symbols must be intuitive, used sparingly (in the figure captions too), and defined at first use. As neither q nor g suggests compliance or non-compliance and are similar in appearance, they are easily confused. Symbol definitions can be repeated in the appendix, along with their values and sources, which however should not repeat other text in the manuscript.
Thank you for your positive and constructive feedback. Thank you for the helpful suggestion. We have revised the paper as suggested by the reviewer. We changed to use the subscript q and nq representing either compliance or non-compliance. The main text was reorganized to make sure all the symbols were explained well when they were first mentioned.

C2. Why distinguish hospitalized and fully isolated if neither contribute to transmission?
Thank you. Indeed, individuals in these compartments do not contribute to the infection, however we assume that deceased cases can only happen in hospital. And this is important because we fitted to death data. We have clarified in our manuscript that fully isolated refers to infection with mild symptoms while hospitalized is severe infection which needs medical care. (Page 4, 15, 18)

C3. Having symptomatic and asymptomatic individuals equally infectious is not consistent with the literature.
Thank you. Although symptomatic and asymptomatic cases might have a different viral load, their transmission might differ in terms of contacts. Asymptomatic cases will maintain their routing without taking any extra precaution, resulting in having more contacts than any other symptomatic will have. We then assume that on average, their transmission is equal. We have clarified it in our revised manuscript. (Page 18)

C4. Why can two members of a family not be infected at the same time?
We are grateful to you for pointing out this issue. In real situations, the number of members in each family in Toronto ranges from 1 to 5 or more, so it is entirely possible for two or more numbers to get infected at same time in a family. The situation that two members of a family can be infected at the same time have been

described in our model. We have deleted this assumption in our revised manuscript. (Page 15)

C5. I suggest using 'effective' versus 'instantaneous' reproduction number throughout unless the authors mean to distinguish empirical- and model-based estimates.
Thank you. We have revised it by using 'effective' reproduction number. (Page 2-7, 12, 25, 28)

C6. I suggest 'onset' versus 'episode'. Onsets are single times, whereas episodes span longer periods.
Thank you for your suggestion. We agree that 'episodes' span longer periods. However, the data we obtained from the Toronto Public Health is using the episode date. And they claim that 'the episode date is a derived variable that best estimates when the disease was acquired and refers to the earliest available date from symptom onset (the first day that COVID-19 symptoms occurred), laboratory specimen collection date (positive result), or reported date'. The data by episode date will update over time. To maintain consistency with our data sources, we prefer using 'episode' in the manuscript and we have clarified the episode date in our revised manuscript. (Page 3, 4)

C7. The word 'since' has a temporal connotation. 'As' or 'because' would be correct in most places. As 'behavior' is a plural noun, 'behaviors' is incorrect. If 'which' is not preceded by a comma, it probably should be 'that'.
Thank you for your suggestion. We have revised it accordingly.

C8. Formulae like that on line 111 are fine in the main text, but ones like that on line 152 belong in the appendix.
Thank you and we have moved the formula on line 152 to the appendix. (Page 25)

C9. I suggest using average versus aggregate numbers in comparisons of results for families differing in size, especially given the curious assumption about only one family member being infected at a time.
Thank you for your suggestion.
We are sorry to confuse you due to the improper expression in the original draft. We have given an explanation and updated the corresponding assumption. Please see the above comment C4 for details.
In order to better understand the severity of the entire epidemic in the structure form of actual epidemic data reported by the Toronto Public Health Bureau, we have retained the original analysis of the aggregate numbers. However, it's a good idea to compare the results for different family size by the average number as suggested. Therefore, in the revised draft, we further added the calculation of the average number of infections by May 6 with average household size of 2 and 3, respectively. We have added the corresponding results in the result section of the revised manuscript with highlights. Please refer to our updated version for details. (Page 5-6)
"With SAHP implementation, if the average household size is 3, by May 6, 2020, there was an average of 11 people infected per 1,000 households, while it was 4 if the average household size is 2."

C10. In the main text, avoid the mathematical shorthand of x (y) ... (e.g., section 3.4).

Thank you for your suggestion. We have deleted the mathematical shorthand of x (y) and revised the results part as suggested by the reviewer. (Page 6)

C11. In the discussion of multiple mitigation measures (line 265), what exactly is meant by measures affecting each other? That their individual effects depend on which others occur simultaneously?

Thank you. From our results, we observe that if we want to keep the number or infections or the reproduction number under a certain threshold, if one of the measures is relaxed, another one needs to be modified accordingly. For example, if we want to keep the risk of infection associated to reopening below 1, then if contacts are increased to 14, the probability of transmission needs to be drastically reduced as well. Further explanation was added to the conclusions. (Page 7)

C12. Use 'rate' for rates and 'proportion' for proportions. Which is Q?

Thank you and we have revised it as suggested. Q is the maximum compliance proportion of SAHP. (Page 16)

Reviewer 2: Affan Shoukat
Institution: Yale University
General comments (author response in bold)

C1. I have reviewed the manuscript "Efficacy of stay-at-home policy and transmission of COVID-19 in Toronto, Canada: a mathematical modeling". In this study, the authors evaluate how stay-at-home policies have affected COVID-19 transmission in Toronto, Canada. In addition, they suggest particular reopening strategies for the city. Their methodology uses a fairly complex SEIR model, with reported case data and deaths for model validation and calibration. The methodology is sound, but I have concerns about some of the assumptions used in the modelling process that cast serious doubts on the validity of the results. I have outlined my concerns below.

Thanks for your appreciation.

C2. One of my major concerns is the calculation of R_t in Section 2.3. The authors assumed a fixed value of 7.5 for the serial interval. Indeed, the serial interval is assumed to be a fixed characteristic of COVID-19 in many modelling studies. However, a recent study, published in Science, suggests that the average serial interval changes as nonpharmaceutical interventions are introduced. For example, the study shows that the mean serial interval of COVID-19 was shortened substantially from 7.8 to 2.6 days within a month of implementing NPI in mainland China. Ideally, the authors should implement a time-dependent function for the serial interval in the equation, although I am not sure how feasible that would be. Alternatively, the authors can run their model with different values of the serial interval (say 2.6 as the mean of the Gamma distribution) and show their results are still robust. If the results (i.e. the effective reproduction number) do vary substantially, then I am afraid the authors will need to include the time component of the serial interval in their R_t calculation. At the very minimum, the study needs to be cited and discussed as a limitation. (Study: https://science.sciencemag.org/content/369/6507/1106).

Thank you. We tried to use different mean values for the gamma distribution of serial interval as suggested by the reviewer. But in the early stages of the epidemic, we thought it was reasonable to use an average of 7.5 days of Gamma distribution for the serial interval, which is also applied in other studies (see

Reference [2]). And despite the extreme mean value of 2.6 days, the difference in R_t is not significant and shows the same trend (Figure S1). Therefore, we believe that our results are relatively robust. In addition, although the real-time serial interval does change with respect to the NPI's implementation, in the time frame used for our work only the stay-at-home policy was enforced. When this measure is relaxed, the period of time is too short and the data available will not suffice to evaluate the real-time serial interval. We understand this is a limitation and it will be further explained in the conclusions. (Page 7)
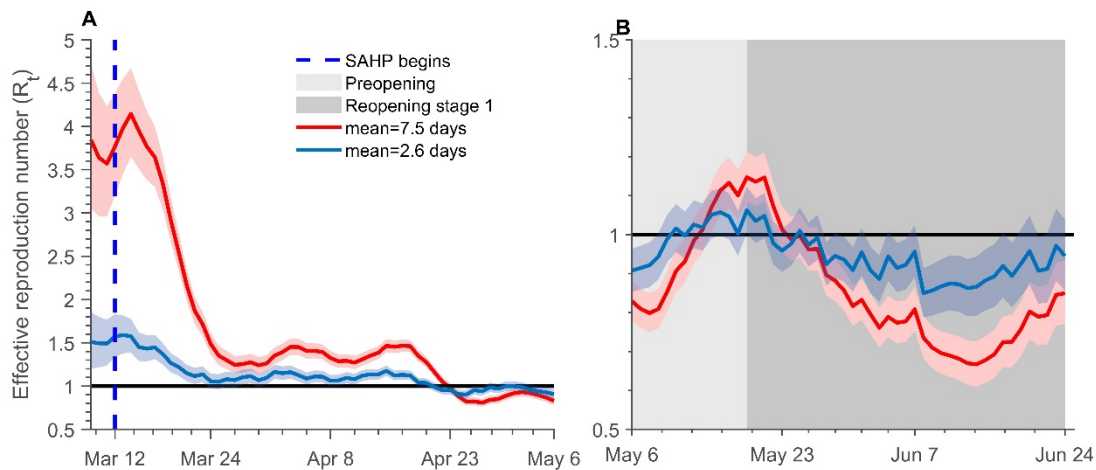
Figure S1: Transmissibility of COVID-19 in Toronto with different mean value of Gamma distribution of serial interval.

C3. The next major flaw for me is the lack of age-structure in the model formulation. The epidemiologic characteristics of COVID-19 are very much dependent on the age-structure of the population. Hospitalization, mild vs severe disease, proportion of asymptomatic infections are all inherently functions of age. While the authors have discussed this is as a limitation, they have not justified why their results are robust without including these details.

Thanks. We agree that the age structure is helpful when model the hospitalization and death from different age groups. The school was closed in our study period (see Reference [3]) and we were looking at community transmission over the time period where kids were not in school, so primary activity was adult work and errands. We focus on the impact of stay-at-home policy on the spread of infection, hence we consider the household as a whole unit. For a household, we ignore the differences between family members and use a well-mixed transmission. The average effect among different age groups can also manifest the epidemic trajectories.

On the other hand, as far as we know that our model is among the early effort to model the household transmission, and our model is a complex one (with many equations) to deal with the spatial complexity of the social structure. If we consider age groups on top of household transmission, it could be too complicated. Therefore, we would rather leave the age structure for future consideration.

Our model has been applied to conduct both short and long-term projections for Toronto Public Health to support the decision making, which shows satisfactory performance. We discussed this as a limitation in our manuscript. (Page 7)

C4. The formula and interpretation for R_{reopen} (Line 150, main text) is not clear. Does this represent a reproduction number? If so, how does it differ from R_t? The formula includes notation/variables that are not defined. Please expand this section to include all the relevant details.

Thanks. The R_{reopen} is a reproduction number while R_t is the effective reproduction number. And R_{reopen} is derived from the model without SAHP when the community is fully reopened. We have included the relevant details in the Appendix B as suggested. (Page 25)

C5. Similarly, the subscripts g, q (used in the Figure 2 model schematic) should be described in the main text. I was initially confused about what these meant until I jumped to the appendix to check model details. Then later found out this is also described in Table 1 (which is not referenced in Section 2.2 at all). Another example of confusion is using the variable Q in the results (Line 184). the table states it is the "Maximum compliance rate induced by SAHP" and that is it an estimated value (based on the model fitting process). However, the results discuss the implications of increasing and decreasing this parameter. For example, it is not clear what the authors mean by saying "If increases from 55% to 75%". Please expand Section 2.2 to include sufficient details about the model; for example, a succinct summary in paragraph form of Table 1 should be included to fully explain the model. This should be done within the scope of journal requirements as well as the target audience.

Thank you, we have included the description of the subscripts in the main text and the caption of the Figure 2. Also, we have revised our method and results part as suggested by reviewer. (Page 3-5, 12)

C6. The overall notation of the mathematics is difficult to follow. Variables and function names seem to be overlapping causing confusion in the interpretation. For example, on Line 44, Page 25, Appendix, what is q_q(t)? Is it the equation defined on line 40? On Line 5, Page 26, Appendix, the variable Q1 is not given an interpretation (not listed in the table of parameters as well). Another example is Line 13, Page27. The authors say the variables i,j,k,l,m,x,y,z denote the number of individuals in each compartment, but there are 8 variables there but only 7 compartments provided. I would recommend the authors carefully reading the mathematical notation and defining all terms and variables. Furthermore, given the medical scope of the journal, the authors should provide a justification or explanation to the equations and variables used so that it is clear for all users.

Thank you for the helpful suggestions. We have revised the manuscript accordingly. $q_q(t)$ is the equation defined on line 40, page 25 and we added the definition of $q_q(t)$. We had the definition of Q_1 in the manuscript, "*We denote the maximum compliance rate (Q_1) as the maximum proportion of the number of people in the group that will carry out SAHP*", and Q (maximum compliance proportion of SAHP) is the estimation of Q_1. In order to distinguish between the two, we used different notations. Also, there are 8 compartments provided in the household and we add the related explanation in our revised manuscript. (Page 20, 22)

C7. Details on the household composition is unclear. The function P(t) is supposedly the number of households with a certain number of susceptible, exposed, infected, etc people. However, the equations suggest to me that all susceptible individuals from all households were homogeneously mixed; similarly all infectious individuals from all households were homogeneously mixed and so on. I may be wrong in my interpretation

here, but a clearer exposition of how household structure was implemented is required.
Thanks. The detailed information about the household composition is given in the section of models in the appendix. The P(t) is the number of certain types of household (eg, two susceptible 1 exposed) and it will change over time due to the dynamic of the epidemic in the household. The individuals in the non-compliance group were homogeneously mixed, and all individuals in a household were homogeneously mixed while the SAHP compliant individuals from all households were not homogeneously mixed. The household following SAHP can only infect their household members and will not transmit the virus to other households. The change rate of individuals in the group ($S_q(t)$, $E_q(t)$, $A_q(t)$, $I_{q1}(t)$, or $I_{q2}(t)$,) that follows SAHP can be calculated by the sum of the change rate of that individuals from all types of households. (Page 18-19, 21-23)

C8. Line 27 of Appendix: "A detailed description of dynamical transmission of COVID-19 is described in the flowchart (Fig. 3)". The reference to the figure is incorrect. The flowchart is shown in Figure 2. Same error on line 30, Page 26.
Thanks. We have revised it accordingly. (Page 18)

C9. The variable `a` in Table 3 is stated to be the proportion of infected with "apparent infection". What is apparent infection? Do the authors mean the proportion of individuals that will develop symptomatic infection? If so, this does not seem correct to me.
Thanks. We have reviewed the definition in the table as prodromal. *a* is the "*proportion of infected with prodromal infection*". (Page 16)

C10. How is the mild cases identified in the symptomatic compartment? Furthermore, on Line 37: "If the quarantine is respected well enough, these infections will be fully isolated and, consequently, will not contribute to the spread of the virus. Otherwise, they are still a source of infection until recovery". So what is the assumption made in the model?
Thanks. I_{nq2}, I_{q2} and *W* are all mild cases with symptoms, while the severe cases are in the H compartment. For mildly symptomatic infected, if they are isolated in the W compartment, they will not infect other people. People in I_{nq2} can infect other people in the non-compliance group, while those in the mild cases in I_{q2} compartment can infect their family members. (Page 4, 15, 18)

C11. In Figure 4, I am assuming that this is the model-based Rt and not based on the reported cases (and that the red band is the uncertainty). This should be clarified in the caption. Alternatively, it would be better in my opinion to include another curve on the plot of the model-free Rc.
Thank you for your comments. The effective reproduction number (R_t) in Figure 4 is calculated based on data by episode date. We have clarified it in the manuscript and revised the caption by adding the detailed information. The R_c is the basic reproduction number, and we may not be able to plot the curve of R_c. To avoiding the confusion, we have renamed R_c to R_0. (Page 12)

Reference
   1. **Kucharski A J, Russell T W, Diamond C, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*, 2020, 20: 553-558.**

   2. **Tang B, Scarabel F, Bragazzi N L, et al. De-Escalation by Reversing the Escalation with a Stronger Synergistic Package of Contact Tracing,**

Quarantine, Isolation and Personal Protection: Feasibility of Preventing a COVID-19 Rebound in Ontario, Canada, as a Case Study. *Biology (Basel)* 2020; 9:100.

3. Toronto. News Releases & Media Advisories. 2020. https://www.toronto.ca/news/.