PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The role of masks, testing and contact tracing in preventing COVID-19 resurgences: a case study from New South Wales, Australia
AUTHORS	Stuart, Robyn M.; Abeysuriya, Romesh; Kerr, Cliff; Mistry, Dina; Klein, Dan; Gray, Richard; Hellard, Margaret; Scott, N

VERSION 1 – REVIEW

REVIEWER	Lewis Spurgin University of East Anglia, UK
REVIEW RETURNED	03-Dec-2020

GENERAL COMMENTS	This is a modelling study on SARS-CoV-2 dynamics parameterised for contact data in New South Wales, Australia. I believe that context specific studies such as this are useful, and from looking at previous work the model itself seems detailed robust. However, I do have a number of major concerns. Firstly, I am concerned about the small number of replicate simulations, and that no estimate of error is provided so that the model output can be assessed. Ideally the number of replicate simulations would be increased, but if this is not possible, then the range needs to presented at the very least. With just 10 or 20 replicates, you could present the individual model runs in the supplementary material. Without this it is quite difficult as a reviewer to assess the model behaviour. Secondly, I think that more details of the original model are needed here. I appreciate that the model is published elsewhere, but a brief reiteration of how e.g. contact patterns and transmission dynamics are simulated seems necessary to me. I also felt that the relationship between testing and tracing could be explained more clearly. For example, in a scenario in which 50% of community contacts are traced and 90% of contacts are tested, how are these tests divided between different types of contacts? Secondly, I think that the highly simplified way in which mask wearing is modelled should be more clearly stated throughout (e.g. by referring to 'reducing transmission efficiency by 30% as may be achieved through mask wearing', rather than just 'mask wearing'). Ideally sensitivity analysis on this parameter would be carried out. Thirdly, a much greater acknowledgement of the wider empirical and modelling literature on contact tracing is required. Numerous modelling studies have shown that contact tracing alone not be sufficient for epidemic control without other measures, and recent empirical work from China support this (Sun et al., 2020, Science).

	On a related note, there are no references in the discussion section at all, and numerous claims are made which require citation. There is discussion about fatigue, complience, vigilance and policy, all of which should be set in the context of the existing literature. I also think that the broader empirical and modelling literature on contact tracing should be discussed in this section, as well as in the discussion, so that work is set within appropriate context. A few specific comments below. I hope these points are helpful, and I wish you the best of luck with your manuscript. p6 Line 45: I'm not convinced that "a balance between masks, testing and contact tracing" is a useful way of thinking about things. Testing generally required for contact tracing, and neither preclude mask wearing. Suggest rephrasing. p8 line 31 - is ten the number of replicate simulations? In the figures it says 20 p8 line 38 - this seems like quite a big assumption, given that it is one of the main aims of the study. Is it worth performing some sensitivity analysis on this?

REVIEWER	Isobel Braithwaite University College London, U.K.
REVIEW RETURNED	15-Dec-2020

GENERAL COMMENTS	I thought this modelling paper was clearly written and well-presented, and addressed an important and policy-relevant set of questions, covering 3 of the main NPIs currently in use to control COVID-19, and evaluating how they interact when used in parallel at different levels. Although the vaccine is now starting to come online, there is likely to be a delay before its full impact can be felt, and access is unlikely to be evenly distributed, so the learning from this modelling study could be useful in many settings as we enter 2021. The introduction and discussion/conclusion probably need some updating to reflect the news re. the vaccine(s) since its submission, and to demonstrate why and how the paper is still relevant. Overall, it would also be helpful if the model could be updated to provide projections for (say) Jan-March or April 2021, for greater policy applicability and ease of communicating the key findings. Specific comments by section: Introduction End of p.5 - it may be worth specifying that the just over 4000 figure is a cumulative (not point) case total for the benefit of readers in countries less familiar with the situation in Australia. Top of p.6 - 'At the same time, high levels of testing (~20,000 tests/day) and rapid contact tracing were in place, with notable focus on contact tracing.' Some additional detail, for example regarding the testing rate per capita relative to other similar/comparable settings and the proportions of cases reached for interview and the proportion of identified contacts reached within a specified timeframe from case diagnosis would be helpful to substantiate this description. In the detail provided below, there is no mention of

	contact tracing of acquaintances or contacts not met in specific venues or who the case can provide details for directly (e.g. family or household members, friends) - is notifying these types of contacts also routinely done? This would be helpful to clarify. Mid page 6 - '45 (5%) of these were locally acquired and not linked to known clusters (20).' Could this be rephrased for clarity? (e.g. '45 (5%) were neither acquired overseas nor linked to a known cluster', if this retains the intended meaning) Methods P.7 - there is a sentence which begins 'We then calibrate the model by adjusting the per-contact transmission rate to fit data on the daily number of cases diagnosed' - I think this should say that the per-contact transmission rate and number of seed infections were both adjusted to calibrate the model? As above, could the model be re-run for a later period (e.g. during 2021)? The Model Analysis section states "To model the efficacy of contact tracing, we assume that 100% of household contacts will be traced and notified on the same day that test results are communicated, and that 95% of school contacts and 90% of workplace contacts will be notified on the following day.' How realistic and reflective of current practice are these assumptions? (They seem high to me, but this may reflect my experience in the UK). Is there any data from the contact tracing system that can substantiate/justify them? (E.g. data on the proportions of contacts known to the case traced within 24 and 48 hours). How were the different proportions of community (venue-related) contacts traceable within a week selected? Is 100% realistic in practice or are 100% and 0% just included for illustrative purposes? (I note that 100% mask-wearing is not assumed and wondered why the difference since I suspect 100% is unattainable for both, albeit for different reasons). Is there any empirical data from tracing efforts in NSW against which this could be benchmarked in the discussion? Regarding mask uptake, I have the impression that a simplifying assumption has been made that an individual who is compliant with masks in the model will do so all of the time when in community settings. However, the relative importance of eating and drinking in social venues (in terms of the overall person-time spent, together with the levels of crowding and ventilation) strikes me as an important factor - as people who are otherwise compliant with mask mandates are also not able to wear masks whilst eating or drinking, e.g. for at least some of the time whilst in a cafe/restaurant/bar/nightclub. Is there a way to either incorporate this complexity in the model (for example using data on the proportion of time spent in different types of hospitality/retail/entertainment venues to estimate the proportion of time within indoor venues spent not wearing masks by otherwise compliant individuals), or to address it in discussion? I would also explicitly mention these types of setting (not jsut arts venues etc) as they appear to be more important venues for transmission, and would make it clearer whether or not masks are assumed to be worn in schools and workplaces in the model, as the the reader's

	interpretation of 'community venues' may differ; perhaps rather than an exhuastive list specifying which settings are excluded may help (e.g. households, ...). It would be good to acknowledge the possibility of varying or low rates of compliance with self-isolation/quarantine requests from contact tracers (or where relayed by household members) more directly within the methods, not only in discussion (and here it would benefit from more in-depth discussion), or at least to better justify the 90% assumption made on this. In the UK this has been found to be low (full compliance with the recommended 14d self-isolation period around 20% and quarantine of contacts even lower- https://www.medrxiv.org/content/10.1101/2020.09.15.20191957v1), though I'm not sure how this compares in Australia - a reference to survey data or similar would be useful if available. I would suspect that the level of compliance is very unlikely to be as high as 90% as detailed later in the same paragraph (though I recognise that the Australian context differs from the UK's in important ways, and there may be data that can justify its choice). In the absence of reliable data, I think a sensitivity analysis of some lower levels of compliance with quarantine advice would strengthen the analysis. This affects both the methods text and Table S1. It may be helpful to specify that the 30% reduction estimate for mask-wearing, based on the IHME reference cited, was adopted by them based on review of data from 2 meta-analyses and one further study, since that reference isn't peer-reviewed itself. Is the 33% (RR 0.67) from their study rounded to 30% in order to be a conservative estimate or for another reason? Secondly, given the 95% C.I. for the relative risk estimated in this report of 0.49-0.88 a sensitivity analysis re. mask efficacy, to reflect this uncertainty, would be useful. Regarding testing - why are symptomatic cases and identified contacts assumed to be tested in the same proportions? It may be helpful to separate out their effects in the analysis, since the drivers of test uptake in the two groups are somewhat different (other than where testing capacity limits are a major factor - but even then only if symptomatic individuals are not prioritised) and their impacts on system dynamics are also likely to differ. On further reading I can see this within a supplementary figure but it does not seem to be clearly referred to in the results section - further discussion of sensitivity analyses would be helpful in the Results. There also appears to be a built-in assumption of 100% sensitivity and specificity in testing, but this is not spelt out (and probably not reflective of available testing options). Exploring the impact of <100% sensitivity and specificity on the findings (and of different levels of test performance, for example relfecting PCR vs lateral flow tests) both on outbreak control and on numbers in quarantine at any one time would be a valuable addition if this is feasible. More generally it may be useful to present the numbers quarantined as in similar studies in the UK (e.g. https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217) the numbers required to be quarantined for R<1 have been modelled as being very large in some scenarios. Results

	I think there's a slight discrepancy as figures 1 and 2 refer to 20 simulations per combination of inputs, but the Methods text refers to 10 being done for each - the more simulation runs that can be done for each permutation of input parameters the better, particularly since only the median run results are presented, but of course this will depend on the processing time involved. It would be good to make it very clear how the results presented (in both the text and figures) are derived from these runs, particularly at the start of the results section. Some further brief description of the uncertainty represented by the stochastic nature of the agent-based model would also be helpful. In the line 'with as little as 70–110 new infections estimated over October 1 – December 31 under high mask uptake scenarios, or 340–1,400 without masks, depending on the efficacy of community contact tracing' Are these the cumulative total new infections across scenarios, with maximum and minimum cumulative values for 75% and 0% mask uptake respectively (and if so can this be specified)? It would be helpful to make clearer exactly what each figure refers to, to enable readers to cross-reference with figure 2. Discussion Overall I thought the discussion was well-written. It struck me that reference could usefully be made in the first part of the Discussion section (paras 2 and/or 3) to the WHO's guiding principles around the choice and prioritisation of NPIs (e.g.'Measures with the highest level of acceptability and feasibility, proven effectiveness–and which minimize the negative consequences on health and well-being of all members of society and the economy–should be considered first.' from https://www.who.int/publications/i/item/considerations-in-adjusting-public-health-and-social-measures-in-the-context-of-covid-19-interim-guidance). The limitations section is generally good, but slightly more detailed discussion on the uncertainties within the model as well as in relation to wider considerations would improve it further. More discussion about the selection of parameters for the 3 interventions of primary interest, how realistic the ranges are and perhaps a clearer indication of which combination of these key parameters the authors think are currently (approximately) the case in NSW would be helpful for interpretation. One of the messages that I thought could have been highlighted more was the importance of asymptomatic testing of contacts, not only testing of symptomatic individuals. It would be good to know more about why this is (probably in the Results section, where figure 3 is referenced), since presumably 90%+ of them are assumed to be quarantining effectively without the testing - is this because it allows the contact tracing and quarantine of their contacts in turn (ie. is the effect only present with reasonable levels of tracing)? Something to explain this, and perhaps an illustration to extend figure 3 for a less optimistic/more realistic contact tracing scenario would be good if possible.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Dr. Lewis G Spurgin, University of East Anglia

No.	Comment	Response
1A	This is a modelling study on SARS-CoV-2 dynamics parameterised for contact data in New South Wales, Australia. I believe that context specific studies such as this are useful, and from looking at previous work the model itself seems detailed robust. However, I do have a number of major concerns.	We sincerely thank the reviewer for taking the time to review, and for providing such thoughtful and constructive comments.
1B	Firstly, I am concerned about the small number of replicate simulations, and that no estimate of error is provided so that the model output can be assessed. Ideally the number of replicate simulations would be increased, but if this is not possible, then the range needs to presented at the very least. With just 10 or 20 replicates, you could present the individual model runs in the supplementary material. Without this it is quite difficult as a reviewer to assess the model behaviour.	We agree with this comment. Since first submitting this manuscript, we have made considerable improvements to both our calibration methodology and our simulation algorithms, which now allow us to run many more simulations. As a result, we have rerun our scenarios with 100 simulations instead of 10, and have also added additional figures to investigate the variance across simulations. Rather than presenting this on a single figure, we have chosen to present the simulation medians in Figure 3, and estimates of the variation across simulations in Figure 4. To translate the variance across simulations into a useful policy metric, we have evaluated the risks associated with various outcomes (i.e., the probability of diagnosing more than a certain number of cases per day, as calculated by looking across all 100 simulations).
1C	Secondly, I think that more details of the original model are needed here. I appreciate that the model is published elsewhere, but a brief reiteration of how e.g. contact patterns and transmission dynamics are simulated seems necessary to me. I also felt that the relationship between testing and tracing could be explained more clearly. For example, in a scenario in which 50% of community contacts are traced and 90% of contacts are tested, how are these tests divided between different types of contacts?	We have added several more paragraphs to the methods section to give detail about how the contact patterns and transmission dynamics were simulated, including a paragraph on contact tracing (3rd methods para, beginning with “Next, we created contact networks for these agents.”), and a description of transmission dynamics (1st methods para, beginning with “Covasim contains detailed descriptions of agedependent disease acquisition and progression probabilities… ”). Since we don’t model a fixed number of tests, we don’t need to stipulate the division of tests between symptomatic people and asymptomatic contacts. To pick up on the example given here: if 50% of community contacts are traced and then 90% of those are tested, these tests do not preclude ongoing testing for symptomatics.

1D	Secondly, I think that the highly simplified way in which mask wearing is modelled should be more clearly stated throughout (e.g. by referring to 'reducing transmission efficiency by 30% as may be achieved through mask wearing', rather than just 'mask wearing'). Ideally sensitivity analysis on this parameter would be carried out.	We have added a sensitivity analysis on the efficacy of masks, as well as additional discussion in the methods section about the difficulties of assessing their “true” efficacy in real-world settings.
1E	Thirdly, a much greater acknowledgement of the wider empirical and modelling literature on contact tracing is required. Numerous modelling studies have shown that contact tracing alone not be sufficient for epidemic control without other measures, and recent empirical work from China support this (Sun et al., 2020, Science).	We have added a number of additional references, especially to the introduction and the discussion.
1F	On a related note, there are no references in the discussion section at all, and numerous claims are made which require citation. There is discussion about fatigue, complience, vigilance and policy, all of which should be set in the context of the existing literature. I also think that the broader empirical and modelling literature on contact tracing should be discussed in this section, as well as in the discussion, so that work is set within appropriate context.	Whilst we don’t have space here to enter a more complete discussion about the literature on contact tracing, we have added mention of it to the introduction (3rd para: “contact tracing means that only those at greatest risk of transmitting the virus need to stay home and have been shown to be effective in numerous settings (15,25–29)”) and to the discussion (4th para: “Although various efforts have been made to synthesise the everexpanding body of research regarding the efficacy of different interventions (1–4,61)”). We have also added more references to the discussion, notably referring to the WHO’s publication on pandemic fatigue.
1G	A few specific comments below. I hope these points are helpful, and I wish you the best of luck with your manuscript.	Again, we thank you for the helpful comments!
1H	p6 Line 45: I'm not convinced that "a balance between masks, testing and contact tracing" is a useful way of thinking about things. Testing generally required for contact tracing, and neither preclude mask wearing. Suggest rephrasing.	We have modified the language used here, to “assess the roles of masks, testing and contact tracing as a means of controlling communitybased transmission”
1I	p8 line 31 - is ten the number of replicate simulations? In the figures it says 20	We have fixed this – all simulations now have 100 replicates

1J	p8 line 38 - this seems like quite a big assumption, given that it is one of the main aims of the study. Is it worth performing some sensitivity analysis on this?	I believe this refers to the assumption that "and that 50% of all other contacts (which we refer to as community contacts) were traced within a week of a case notification, with a mean time to trace of one day". We do indeed perform sensitivity analysis on this parameter, with values of 0-100% considered in the scenarios.

Reviewer 2: Dr. Isobel Braithwaite, UCL

No.	**Comment**	**Response**
2A	I thought this modelling paper was clearly written and wellpresented, and addressed an important and policy-relevant set of questions, covering 3 of the main NPIs currently in use to control COVID-19, and evaluating how they interact when used in parallel at different levels.	Our sincere thanks to the reviewer for providing such a thoughtful and detailed review!
2B	Although the vaccine is now starting to come online, there is likely to be a delay before its full impact can be felt, and access is unlikely to be evenly distributed, so the learning from this modelling study could be useful in many settings as we enter 2021. The introduction and discussion/conclusion probably need some updating to reflect the news re. the vaccine(s) since its submission, and to demonstrate why and how the paper is still relevant. Overall, it would also be helpful if the model could be updated to provide projections for (say) Jan-March or April 2021, for greater policy applicability and ease of communicating the key findings.	We have modified the language throughout, especially in the introduction and discussion, to reflect the current epidemiological conditions, including the vaccine and emergence of new strains. As data is now available, we are also now able to compare the model projections to what eventuated over October-December 2020, which we believe serves as a useful validation for the model and one that is rarely undertaken in modelling studies However, we believe that it is still best for us to focus on the October-December period, rather than providing projections for Jan-March. The main reason for this is that this piece of work was important for the NSW government as part of their assessment of whether to make masks mandatory. As it turned out, masks were eventually made mandatory in the state, but not until after a new outbreak emerged towards the end of 2020. This outbreak necessitated burdensome new restrictions on interstate travel and the sizes of gatherings immediately before Christmas and New Years Eve. Such an outbreak was predicted by the model as likely to occur if testing rates declined. So we believe that it is valuable to show that these outcomes were known to be likely back in September. We have added points to the discussion section to address this.

2C	Introduction End of p.5 - it may be worth specifying that the just over 4000 figure is a cumulative (not point) case total for the benefit of readers in countries less familiar with the situation in Australia.	We have modified the sentence to clarify that we mean cumulative cases here.
2D	Top of p.6 - 'At the same time, high levels of testing (~20,000 tests/day) and rapid contact	We have added data on testing rate per capita, along with a comparison to other countries (5th para of the intro: "At the same time, high levels of testing were in place, with ~20,000

	tracing were in place, with notable focus on contact tracing.' Some additional detail, for example regarding the testing rate per capita relative to other similar/comparable settings and the proportions of cases reached for interview and the proportion of identified contacts reached within a specified timeframe from case diagnosis would be helpful to substantiate this description. In the detail provided below, there is no mention of contact tracing of acquaintances or contacts not met in specific venues or who the case can provide details for directly (e.g. family or household members, friends) - is notifying these types of contacts also routinely done? This would be helpful to clarify.	people tested per day over June–September (~2.7/day per 1,000 people), resulting in an average testing yield of 0.05%, one of the world's lowest (39)" We have added detail about NSW Health's contact tracing protocol this in the methods section: "Relative to many other contact tracing programs across the world, the NSW contact tracing program was differentiated by its extensive efforts to identify a person's community contacts in addition to their household, social, school, and workplace contacts (42)"
2E	Mid page 6 - '45 (5%) of these were locally acquired and not linked to known clusters (20).' Could this be rephrased for clarity? (e.g. '45 (5%) were neither acquired overseas nor linked to a known cluster', if this retains the intended meaning)	We have modified as suggested, to: "Over the four months from June 1 – September 30, 2020, ~900 new cases were identified, but only 45 (5%) of were classified as "source unknown", meaning that they were neither acquired overseas/interstate nor linked to known clusters (41)"

2F	Methods P.7 - there is a sentence which begins 'We then calibrate the model by adjusting the per-contact transmission rate to fit data on the daily number of cases diagnosed' - I think this should say that the per-contact transmission rate and number of seed infections were both adjusted to calibrate the model?	Yes, we have modified.
2G	As above, could the model be rerun for a later period (e.g. during 2021)?	Please see our response to comment 2B explaining why we think it is best to leave this as-is.
2H	The Model Analysis section states "To model the efficacy of contact tracing, we assume that	We recognise that these values are high in a global context. However, available reports do seem to justify these assumptions. We have elaborated and added sources in the

	100% of household contacts will be traced and notified on the same day that test results are communicated, and that 95% of school contacts and 90% of workplace contacts will be notified on the following day.' How realistic and reflective of current practice are these assumptions? (They seem high to me, but this may reflect my experience in the UK). Is there any data from the contact tracing system that can substantiate/justify them? (E.g. data on the proportions of contacts known to the case traced within 24 and 48 hours).	methods, as follows: “To model the efficacy of contact tracing, we use publicly-available data from NSW Health (42). We note that, although the program reports the proportion of known contacts that were reached within defined timeframes, we would ideally like to know the proportion of all contacts that were reached, which will be lower than the reported values since it will also include contacts that the case did not recall or disclose. Thus, although NSW Health’s published reports (42) indicate that 100% of contacts are notified within 48 hours, we use slightly more conservative values, namely that 100% of household contacts will be traced and notified on the same day that test results are communicated, and that 95% of school contacts and 90% of workplace contacts will be notified on the following day.”

2I	How were the different proportions of community (venue-related) contacts traceable within a week selected? Is 100% realistic in practice or are 100% and 0% just included for illustrative purposes? (I note that 100% mask-wearing is not assumed and wondered why the difference since I suspect 100% is unattainable for both, albeit for different reasons). Is there any empirical data from tracing efforts in NSW against which this could be benchmarked in the discussion?	We included scenarios looking at 100% contact tracing because this is in line with historical success rates (see e.g. https://www.health.gov.au/resources/collections/coronaviruscovid-19-common-operating-picture, also cited in the manuscript) However, for mask usage, we did not model 100% mask wearing as it is not possible to wear masks 100% of the time, for example since we're modelling contexts like cafes and restaurants where people need to remove their masks to dine. So 75% seemed like a reasonable upper threshold for the mask scenarios – noting of course that it is possible to infer the results for other values by looking at the values that we have modelled.
2J	Regarding mask uptake, I have the impression that a simplifying assumption has been made that an individual who is compliant with masks in the model will do so all of the time when in community settings. However, the relative importance of eating and drinking in social venues (in terms of the overall person-time spent, together with the levels of crowding and ventilation) strikes me as an important factor - as people who are otherwise compliant with mask mandates are also not able to wear masks	This is a good point, and one that we attempted to make some allowance for by only modelling mask uptake of up to 75%, not 100% (as noted in the response to the previous comment). To address this further, we have: • added to the discussion, noting the limitations of not being able to differentiate between mask usage in restaurants vs other settings (see limitations para) • added details to the methods section to specifically outline which settings were included as "community venues" (see para beginning: "Next, we created contact networks for these agents")

	whilst eating or drinking, e.g. for at least some of the time whilst in a cafe/restaurant/bar/nightclub. Is there a way to either incorporate this complexity in the model (for example using data on the proportion of time spent in different types of hospitality/retail/entertainment venues to estimate the proportion of time within indoor venues spent not wearing masks by otherwise compliant individuals), or to address it in discussion? I would also explicitly mention these types of setting (not jsut arts venues etc) as they appear to be more important venues for transmission, and would make it clearer whether or not masks are assumed to be worn in schools and workplaces in the model, as the the reader's interpretation of 'community venues' may differ; perhaps rather than an exhuastive list specifying which settings are excluded may help (e.g. households, ...).	
2K	It would be good to acknowledge the possibility of varying or low rates of compliance with selfisolation/quarantine requests from contact tracers (or where relayed by household members) more directly within the methods, not only in discussion (and here it would benefit from more in-depth discussion), or at least to better justify the 90% assumption made on this. In the UK this has been found to be low (full compliance with the recommended 14d self-isolation period around 20% and quarantine of contacts even lower (ref), though I'm not sure how this compares in Australia - a reference to survey data or similar would be useful if available. I would suspect that	As with our response to comment 2H, we recognise that these values are high in a global context, and have noted as such in the manuscript, but we believe these assumptions are reasonable in this context. See para: "Finally, we assume that people who have been contact-traced will quarantine with 90% compliance from their workplace, school, and community contacts. Whilst this assumption may be optimistic in other global contexts, the lower case counts in NSW mean that contact tracers have far greater capacity for rigorous ongoing follow-up of contacts, and breaches of isolation are escalated with local authorities, as stipulated in national guideline documents (50)."). We would like to perform sensitivity analysis on this, but with 80 core scenarios and 160 more presented in the supplementaries, we fear that the manuscript will get too difficult to understand.

	the level of compliance is very unlikely to be as high as 90% as detailed later in the same paragraph (though I recognise that the Australian context differs from the UK's in important ways, and there may be data that can justify its choice). In the absence of reliable data, I think a sensitivity analysis of some lower levels of compliance with quarantine advice would strengthen the analysis. This affects both the methods text and Table S1.	
2L	It may be helpful to specify that the 30% reduction estimate for mask-wearing, based on the IHME reference cited, was adopted by them based on review of data from 2 meta-analyses and one further study, since that reference isn't peer-reviewed itself. Is the 33% (RR 0.67) from their study rounded to 30% in order to be a conservative estimate or for another reason? Secondly, given the 95% C.I. for the relative risk estimated in this report of 0.49-0.88 a sensitivity analysis re. mask efficacy, to reflect this uncertainty, would be useful.	We have added more references around mask efficacy, and added a sensitivity analysis around mask efficacy.

2M	Regarding testing - why are symptomatic cases and identified contacts assumed to be tested in the same proportions? It may be helpful to separate out their effects in the analysis, since the drivers of test uptake in the two groups are somewhat different (other than where testing capacity limits are a major factor - but even then only if symptomatic individuals are not prioritised) and their impacts on system dynamics are also likely to differ. On further reading I can see this within a supplementary figure but it does not seem to be clearly referred to in the results section - further discussion of sensitivity analyses would be helpful in the Results.	We have integrated the results from the sensitivity analyses more into the main text.
2N	There also appears to be a builtin assumption of 100% sensitivity and specificity in testing, but this is not spelt out (and probably not reflective of available testing options). Exploring the impact of <100% sensitivity and specificity on the findings (and of different levels of test performance, for example relfecting PCR vs lateral flow tests) both on outbreak control and on numbers in quarantine at any one time would be a valuable addition if this is feasible. More generally it may be useful to present the numbers quarantined as in similar studies in the UK (e.g. ref) the numbers required to be quarantined for R<1 have been modelled as being very large in some scenarios.	We agree it would be useful to add results on the number quarantined, and have added this: see 1st para of results: "We further estimate that 79,330 (60,100–129,020) people had been required to self-isolate at some point by September 30, 2020 as a result of having potentially been in contact with a confirmed case." Regarding test sensitivity: There is a parameter for test sensitivity in the model, but we accidentally omitted from the parameters table S1, so we have now added a description as follows: "Our assumptions around test sensitivity coupled with our modelling of viral load kinetics are specified such that the probability of identifying a true positive increases and then decreases over the course of an infection, following a similar profile to that reported in the literature (9–10)." .

2O	Results I think there's a slight discrepancy as figures 1 and 2 refer to 20 simulations per combination of inputs, but the Methods text refers to 10 being done for each - the more simulation runs that can be done for each permutation of input parameters the better, particularly since only the median run results are presented, but of course this will depend on the processing time involved. It would be good to make it very clear how the results presented (in both the text and figures) are derived from these runs, particularly at the start of the results section. Some further brief description of the uncertainty represented by the stochastic nature of the agentbased model would also be helpful.	As noted in our response to comment 1B, we have made considerable improvements to both our calibration methodology and our simulation algorithms, which now allow us to run many more simulations. As a result, we have rerun our scenarios with 100 simulations instead of 10, and have also added additional figures to investigate the variance across simulations. Rather than presenting this on a single figure, we have chosen to present the simulation medians in Figure 3, and estimates indicating the degree of variation across simulations in Figure 4.

2P	In the line 'with as little as 70–110 new infections estimated over October 1 – December 31 under high mask uptake scenarios, or 340–1,400 without masks, depending on the efficacy of community contact tracing' . Are these the cumulative total new infections across scenarios, with maximum and minimum cumulative values for 75% and 0% mask uptake respectively (and if so can this be specified)? It would be helpful to make clearer exactly what each figure refers to, to enable readers to crossreference with figure 2.	Yes, this is what was meant, and we have modified the text to make this clearer, as follows: “all strategies in which there is at least some contact tracing in place and testing rates are very high … lead to a robustly controlled epidemic, with **a median** of ~180 infections in total estimated over October 1 – December 31 under high mask uptake scenarios, or 260–1,200 without masks.”

2Q	Discussion Overall I thought the discussion was well-written. It struck me that reference could usefully be made in the first part of the Discussion section (paras 2 and/or 3) to the WHO's guiding principles around the choice and prioritisation of NPIs (e.g.'Measures with the highest level of acceptability and feasibility, proven effectiveness– and which minimize the negative consequences on health and wellbeing of all members of society and the economy– should be considered first.' (from ref).	We have added a number of additional references, especially to the introduction and the discussion, including the one suggested here - thank you for the suggestion.
2R	The limitations section is generally good, but slightly more detailed discussion on the uncertainties within the model as well as in relation to wider considerations would improve it further. More discussion about the selection of parameters for the 3 interventions of primary interest, how realistic the ranges are and perhaps a clearer indication of which combination of these key parameters the authors think are currently (approximately) the case in NSW would be helpful for interpretation.	We have added quite a bit of detail to the methods section to highlight the way the various model parameters were chosen, as described in the responses to comments 2H-K.

2S	One of the messages that I thought could have been highlighted more was the importance of asymptomatic testing of contacts, not only testing of symptomatic individuals. It would be good to know more about why this is (probably in the Results section, where figure 3 is referenced), since presumably 90%+ of them are assumed to be quarantining effectively without the testing - is this because it allows the contact tracing and quarantine of their contacts in turn (ie. is the effect only present with reasonable levels of tracing)? Something to explain this, and perhaps an illustration to extend figure 3 for a less optimistic/more realistic contact tracing scenario would be good if possible.	We have elected not to show more pessimistic contact tracing scenarios, as it would be contrary to all information provided to us by NSW Health, and indeed, evidence seems to support the fact that the NSW contact tracing program has been extremely effective to date. However, we have highlighted the importance of asymptomatic contact tracing more in the discussion, emphasising the very good point raised here that it enables more thorough contact tracing: "Since our core analyses already assume high compliance with recommended self-isolation policies for known contacts of confirmed cases, the marginal benefit of high asymptomatic contact tracing is primarily to further bolster the efficacy of contact tracing, since it allows for the identification of chains of transmission even in the absence of symptoms."

VERSION 2 – REVIEW

REVIEWER	Braithwaite, Isobel UCL, Institute for Health Informatics
REVIEW RETURNED	04-Mar-2021

GENERAL COMMENTS	The paper is very clear, timely and important for policy decisions, and has definitely improved following the revisions made. A few specific comments below which are comments or minor suggested edits that I think could improve it further; however none of these changes are essential as it's already very good. Abstract The second sentence in the Abstract results is quite long/hard to follow, and it might be helpful to specify 260-1,200 infections near the end of it for clarity. I wonder if the final sentence of the Results might be more intuitive if the probability of this outcome under a median-case scenario is given instead? Particularly as in international context 254 is still very low. It looks from the introduction as though mask use is known to have been generally lower than that in the scenario you're comparing with here, so maybe worth stating that as a likely contributor to the discrepancy if keeping the current framing. Strengths and limitations summary – bullet point 2: suggest 'estimate' rather than quantify the probability as quantify may suggest empirical measurement rather than a modelled estimate.

	Introduction Minor – suggest policy contexts rather than settings at the end of para 1 The wording in lines 30-32 (p.5) may suggest to some readers that the strategies are each effective in isolation – which arguably they are but potentially not enough to bring R<1 if that is the aim. ? reword. Line 6 p.6 – the 4700 figure seems at odds with the one in the abstract – possibly helpful to specify how many were travel-related vs. locally-acquired. p.7 lines 18-19 – I would argue that some of the factors influencing adherence to/support for self-isolation, access to testing et are within policy control (at least somewhat) – potentially rephrase to avoid suggesting these are not relevant areas for public health intervention. Lines 23-25 (end of intro) – perhaps make clear it's the interactions/synergies between these interventions that are of particular interest/ Methods I think the changes made, and the updated timeframes, have really strengthened the paper. p.8 lines 1-7 A brief summary of key modifiable parameters and the values for them selected in the model, perhaps a table in supplementary material, may be useful for readers. (As an example I was looking for the asymptomatic:symptomatic case ratio that had been specified, and the asymptomatic cases' relative infectiousness, since this interacts with testing regimes and is relevant to interpreting the results, but couldn't find this). p.10 line 22-24 – The evidence for non-medical masks, particularly cloth ones, seems to suggest that the benefit is particularly related to source control, though there may be some, more limited, protection for wearers too (as outlined by CDC for example https://www.cdc.gov/coronavirus/2019-ncov/more/masking-science-sars-cov2.html). It would be helpful to make clear whether this directionality is reflected in the model e.g. is this risk reduction assumed to be the same (in a given contact event) whether one or both parties are wearing masks, or is it specifically regarding the 'mask-wearer' agent's risk? p.11 lines 11-14 Very helpful to have this local context. In the scenarios, I'm not sure if I'm correct but I think the base case assumption has been changed to have asymptomatic testing rates equal to symptomatic rates, when it was previously assumed to be lower. What's the rationale for this? As I'm not sure this is the case in most of the world – but perhaps it has been in Australia? If so it may be helpful to clarify, though perhaps limits generalisability as in general I think a symptom-related discrepancy in testing rates (and probably longer delays for asymptomatic testing) has been quite widely observed. I think this refers specifically to asymptomatic contacts of known cases (not to other asymptomatic cases who aren't identified as contacts, say) – if correct, it may be useful to spell this out. Results

	p.13 Is there an interaction between contact tracing effectiveness and the proportion (and timeliness) of testing of asymptomatic contacts of cases? I.e. if contact tracing is limited/slow, is it realistic to expect early enough testing of the asymptomatic contacts independently of this? (I'm assuming contact tracing is what prompts many/some of them to get tested). Perhaps the interlinked nature of contact tracing and testing is a point for some brief discussion, since they are treated as separate in the modelling, e.g. around the top of p.16. I'm not sure the impact of high levels of testing of asymptomatic contacts is spelt out specifically in the Results, though I may have missed it, but it's referred to in the discussion in relation to the added benefit it brings to contact tracing efforts, so it would be good to draw out clearly. If there is a way to provide an indicative/illustrative quantitative example of the difference between 75% mask wearing with 15% and 45% effectiveness in (say) a middle-of-the-road tracing and testing scenario, in terms of the median number of infections expected perhaps, I think this may be very useful for policy-makers thinking about what kind of quality of masks to require/other policies such as vouchers for higher-quality ones as is starting to happen in Europe. Discussion No further comments beyond those made above which relate also to the Discussion – it reads well and makes some important points, and covers several key limitations.

VERSION 2 – AUTHOR RESPONSE

Response to 2nd round review: "The role of masks, testing and contact tracing in preventing COVID-19 resurgences: a case study from New South Wales, Australia"

Reviewer 2: Dr. Isobel Braithwaite, UCL

No.	Comment	Response
2A	The paper is very clear, timely and important for policy decisions, and has definitely improved following the revisions made. A few specific comments below which are comments or minor suggested edits that I think could improve it further; however none of these changes are essential as it's already very good.	Many thanks for such a thorough and helpful review!
2B	Abstract The second sentence in the Abstract results is quite long/hard to follow, and it might be helpful to specify 260-1,200 infections near the end of	We have restructured the sentence a little and moved 260-1200 infections to the end. We have also reworded the final sentence of the Results section of the Abstract to highlight the probability of diagnosing

	it for clarity. I wonder if the final sentence of the Results might be more intuitive if the probability of this outcome under a median-case scenario is given instead? Particularly as in international context 254 is still very low. It looks from the introduction as though mask use is known to have been generally lower than that in the scenario you're comparing with here, so maybe worth stating that as a likely contributor to the discrepancy if keeping the current framing.	254 cases under the conditions that appear to have been in place in NSW over this period.
2C	Strengths and limitations summary – bullet point 2: suggest 'estimate' rather than quantify the probability as quantify may suggest empirical measurement rather than a modelled estimate.	We have changed this as suggested.
2D	Introduction Minor – suggest policy contexts rather than settings at the end of para 1	We have changed this as suggested.
2E	The wording in lines 30-32 (p.5) may suggest to some readers that the strategies are each effective in isolation – which arguably they are but potentially not enough to bring R<1 if that is the aim. ? reword.	We have reworded to clarify this: "When used in combination with physical distancing and hand-washing/hygiene measures, these three strategies together can allow relatively high mobility"
2F	Line 6 p.6 – the 4700 figure seems at odds with the one in the abstract – possibly helpful to specify how many were travel-related vs. locally-acquired.	We have specified what % of these cases were travel-related: "The context for our study is the Australian state of New South Wales (NSW), with a population of 7.5 million and a cumulative total of just over 4700 diagnosed cases as of December 31, 2020, of which the majority (~57%) were acquired overseas"
2G	p.7 lines 18-19 – I would argue that some of the factors influencing adherence to/support for self-isolation, access to testing et are within policy control (at least somewhat) – potentially rephrase to avoid suggesting these are not relevant areas for public health intervention.	We have removed this statement as we agree that it was ambiguous.

2H	Lines 23-25 (end of intro) – perhaps make clear it's the interactions/synergies between these interventions that are of particular interest/	We have added this: In this study, we consider a range of testing and contact tracing levels, and assess the interacting roles of masks, testing and contact tracing as a means of controlling community-based transmission.
2I	Methods I think the changes made, and the updated timeframes, have really strengthened the paper.	Thank you for the suggestions!
2J	p.8 lines 1-7 A brief summary of key modifiable parameters and the values for them selected in the model, perhaps a table in supplementary material, may be useful for readers. (As an example I was looking for the asymptomatic:symptomatic case ratio that had been specified, and the asymptomatic cases' relative infectiousness, since this interacts with testing regimes and is relevant to interpreting the results, but couldn't find this).	We have put in a reference to the specific tables where these parameter values can be accessed: Covasim contains detailed descriptions of age-dependent disease acquisition and progression probabilities and the duration of disease by acuity (Tables 1-2 of (45))
2K	p.10 line 22-24 – The evidence for non-medical masks, particularly cloth ones, seems to suggest that the benefit is particularly related to source control, though there may be some, more limited, protection for wearers too (as outlined by CDC for example ref). It would be helpful to make clear whether this directionality is reflected in the model e.g. is this risk reduction assumed to be the same (in a given contact event) whether one or both parties are wearing masks, or is it specifically regarding the 'mask-wearer' agent's risk?	We have specified that we are just using a single value as an average across different permutations of one or both people wearing masks: "We do not model the differences between both people wearing masks vs source-only vs target only, so the estimates we use here should be considered as averages across these different possibilities."
2L	p.11 lines 11-14 Very helpful to have this local context.	Excellent!
2M	In the scenarios, I'm not sure if I'm correct but I think the base case assumption has been changed to have asymptomatic testing rates equal to symptomatic rates, when it was previously assumed to be lower.	This is as it was before, and seems consistent with the available data from NSW, as well as with the state of epidemic control that has been achieved there. As noted, this refers to asymptomatic contacts of those who've been notified as a contact, not asymptomatics in general - we have clarified this by changing the

	What's the rationale for this? As I'm not sure this is the case in most of the world – but perhaps it has been in Australia? If so it may be helpful to clarify, though perhaps limits generalisability as in general I think a symptom-related discrepancy in testing rates (and probably longer delays for asymptomatic testing) has been quite widely observed. I think this refers specifically to asymptomatic contacts of known cases (not to other asymptomatic cases who aren't identified as contacts, say) – if correct, it may be useful to spell this out.	labelling in Table 1 to "Asymptomatic testing of known contacts"
2N	Results p.13 Is there an interaction between contact tracing effectiveness and the proportion (and timeliness) of testing of asymptomatic contacts of cases? I.e. if contact tracing is limited/slow, is it realistic to expect early enough testing of the asymptomatic contacts independently of this? (I'm assuming contact tracing is what prompts many/some of them to get tested). Perhaps the interlinked nature of contact tracing and testing is a point for some brief discussion, since they are treated as separate in the modelling, e.g. around the top of p.16.	We have added a sentence on this to the discussion: "The interdependencies between different arms of testing and tracing strategies are complex, as testing of asymptomatic contacts is only feasible if contact tracing is effective enough to identify them."
2O	I'm not sure the impact of high levels of testing of asymptomatic contacts is spelt out specifically in the Results, though I may have missed it, but it's referred to in the discussion in relation to the added benefit it brings to contact tracing efforts, so it would be good to draw out clearly.	This can be found in the section on the sensitivity analyse: "Within this set of scenarios, the total number of infections is estimated to be around 50% higher if asymptomatic contacts test at a lower rate than people with symptoms (averaged across all levels of mask usage)"
2P	If there is a way to provide an indicative/illustrative quantitative example of the difference between 75% mask wearing with 15% and 45% effectiveness in (say) a middle-of-the-road tracing and testing scenario, in terms of the median number of infections expected perhaps, I think this may be very useful for policy-makers thinking about what kind of quality of masks to require/other policies such as	We have added this to the end of the Results section: "For example, we estimate 1500 infections over the three-month period assuming 50% tracing, 65% symptomatic testing and 75% mask uptake if masks have an efficacy of 30% (Figure 3, central panel), compared to 1800 infections if masks had 15% efficacy or 1300 infections if masks had 45% efficacy."

	vouchers for higher-quality ones as is starting to happen in Europe.	
2Q	Discussion No further comments beyond those made above which relate also to the Discussion – it reads well and makes some important points, and covers several key limitations.	Thank you!