PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Secondary attack rate of COVID-19 among contacts and risk factors, Tamil Nadu, March-May 2020: A retrospective Cohort Study
AUTHORS	Karumanagoundar, Kolandaswamy; Raju, Mohankumar; Ponnaiah, Manickam; Kaur, P; viswanathan, Vidhya; Rubeshkumar, Polani; Sakthivel, Manikandanesan; Shanmugiah, Porchelvan; Ganeshkumar, Parasuraman; Muthusamy, Santhosh Kumar; Sendhilkumar, Muthappan; Venkatasamy, Vettrichelvan; Sambath, Irene; Ilangovan, Kumaravel; M, Jagadeesan; Govindarajan, Rameshkumar; Shanmugam, Soundammal; Rajarathinam, Selvakumar; Suresh, KST; Varadharajan, M; Thiagarajan, Manivannan; Jagadeeshkumar, K; Ganesh, Velmurugan; Kumar, Sateesh; Venkatesan, Prakash; Nallathambi, Yogananth; Palani, Sampath; Selvavinayagam, TS; Reddy, Madhusudhan; Rajesh, Beela; Murhekar, Manoj

VERSION 1 – REVIEW

REVIEWER	Shah, Komal Public Health Foundation of India, Indian Institute of Public Health, Gandhinagar
REVIEW RETURNED	10-Apr-2021

GENERAL COMMENTS	The paper adequately address the research question and can be accepted in current form.

REVIEWER	Wilkinson, Krista University of Manitoba, Vaccine and Drug Evaluation Centre
REVIEW RETURNED	12-Apr-2021

GENERAL COMMENTS	Thank you for inviting me to review the manuscript Secondary attack rate of COVID-19 among contacts and risk factors, Tamil Nadu, April-May 2020: A retrospective cohort study. This manuscript contains sufficient detail overall, however some clarity is required with respect to the study methods and interpretation. I've separated my review into major and minor comments. Major comments: 1. Quite a few analyses of SARS-CoV-2 secondary attack rates have previously been published, including a recent systematic review and meta-analysis. In the discussion section, the authors should place their findings within the larger body of research and highlight what additional value is provided by their analysis.

2. I would suggest that the authors do not present collapsed results for the community and household contacts and explore whether these results are subject to effect measure modification. Given that some districts limited contact tracing to household contacts only, and that testing was prioritized for household contacts of symptomatic primary cases there is evidence that a head-to-head comparison such as in the risk factors section may be inappropriate. I strongly recommend reconsidering the analyses as presented in Tables 4 and 5 where the reference category is community contacts of non-congregation primary cases. The descriptive results showed that ¾ of the available contacts were community contacts, although very few became cases. The authors further state in the discussion that there was limited interaction at the community level, which also supports the need to present separately.

3. Who were the asymptomatic primary cases? Why were they tested/how were they identified if primary cases were defined as having no established contact history?

4. One of the main findings of this paper was that transmission was higher among contacts of primary cases exposed to the congregation. What hypotheses do the authors propose to explain why the secondary attack rate was higher in these contacts? A quarter of the primary cases were linked to this congregation which occurred early in the local epidemic (March 21-23). Some clarification is needed around the timing of the congregation exposure and the implementation of strict lockdown measures. Did the earlier (congregation) cases occur during a period of less strict public health measures? Did the median number of household contacts change over the study period?

Minor comments:

1. Abstract - please clarify what you mean by congregation exposure as it only becomes clear when the body of the article is read.

2. Strengths bullets - bullet two states that the "Majority of the contacts were tested with RT-PCR, therefore the estimates of secondary attack rate were reliable". Consider rewording; the SAR estimates were based on other assumptions that might make these estimates less reliable

3. Methods –

a. Study setting and the COVID-19 context – please clarify what you mean by the nine administrative districts reporting maximum cases during the study period. Does this mean the analysis is focused on the nine districts with the highest case count?

b. When were contacts tested relative to their quarantine period? e.g. at day 5 after last exposure to a case. Please present the median and IQR in the results

c. How was date of the contact's last exposure to the case determined? Please present the median and IQR and/or minimum and maximum periods between last exposure to the primary case and onset of symptoms/date of diagnosis for the secondary cases

d. Sampling and sample size – please clarify the study period. The study period in the abstract was March 14 to May 5, in this methods section was March 1 to May 30, and tables and title say April -May.

e. Operation definitions - More details are required for these definitions

i. Primary case – does this group represent confirmed cases not named as contacts and without a relevant community exposure? Did you use symptom onset date (or date of testing for

asymptomatic individuals) when determining who was the primary case in a household? How did you deal with co-primary cases (i.e. household contacts with symptom onset date within one day of primary case)?

ii. Contact – please clarify definition of "exposed", e.g. unprotected contact within 2 metres for 15 or more minutes with a confirmed case

iii. Household contact – how did you determine individuals within a household? How did you determine that contact occurred and date of last exposure to case?

iv. Cluster – what would be considered an unusual aggregation in this setting?

v. Secondary attack rate – please clarify this definition; I was unsure what you meant by minus the primary cases of the contacts

4. Results -

a. Description of cases and contacts – I don't understand the sentence about duration of data abstraction and am unsure what value these data add to the manuscript; suggest removal

b. Secondary attack rates - please include confidence intervals in all tables

c. Possible typo – last sentence states, "of the 598 contacts who tested positive"; should this be 599?

d. Previous research has indicated that SARS-CoV-2 tends towards over-dispersion. How many of your primary cases did not transmit within their households?

e. Table 1. I would suggest that the ecological columns of population density and family size could be deleted in this table as well as the days since first case column. I would be more interested to see the contacts columns split into household and community columns

f. Table 2 – as with Table 1, I would recommend presenting columns for both the community and household contact types. Please include confidence intervals.

g. Did SAR vary by district?

h. Tables 4 and 5, please see major comment 2

REVIEWER	Marks, Michael London School of Hygiene and Tropical Medicine, Clinical Research Department
REVIEW RETURNED	19-Jul-2021

| GENERAL COMMENTS | 1) What PCR target was used - single gene/multiple - which genes?
2) What time were individuals visited, were they tested only once? This could significantly impact on the accuracy of PCR testing of contacts - i.e missing people if tested too early or too late which could result in biased estimates. More information on how the contact tracing and testing was performed is therefore needed.
3) Do you have data on approximate viral load (either by quantification or through the CT value of the PCR). Several previous papers have shown a relationship between viral load and risk of onward transmission and so it would be interesting/useful to know if similar relationships were seen here. See for example
a) https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30985-3/fulltext
b)http://modmedmicro.nsms.ox.ac.uk/wp-content/uploads/2021/01/infectivity_manuscript_20210119_merged.pdf
4) How was symptomatic disease defined - please include this in the methods.
5) Operationally how was a cluster defined? |

	6) Operationally how was a congregation exposure defined?
	Minor
	Page 9 - dont mix harvard and vancouver style referencing (line 50)

REVIEWER	Bhatia, Rajiv Stanford University
REVIEW RETURNED	21-Jul-2021

GENERAL COMMENTS	Comments to authors
	1. There are multiple references and analyses of risks in relation to "congregation exposure;" however, "congregation exposure" term is never explicitly and operationally defined in the methods. There is an oblique discussion in the introduction to the paper about a cluster associated with a religious congregation in New Delhi that began on February 9, 2021. Presumably, one assumes that "congregation exposure" refers to this particular religious congregation; however, the term congregation exposure could many have other meanings (i.e., any gatherings) and should be more precisely defined. The source of information for 'congregation exposure in the primary case" should also be precisely described given its import in the article.
	2. The data for this analysis comes from government public health case identification and contact tracing efforts. Authors should summarize written protocols for case identification, contact identification, and contact follow up used by these government agencies and ideally provide as supplementary materials If the protocols varied by jurisdiction, their salient aspects could be synthesized in a table organized by jurisdiction. Salient information in the protocols includes the criteria for contact (e.g., time relative to the case identification data, duration of contact, proximity), period and duration of contact follow-up following contact identification, and timing of symptom and biological disease assessments relative to the timing of contact indentification.
	3. Most "contacts" were non-household. It would be very useful given the paucity of scholarship on exposure risks among non-household community and workplace contacts to estimate SAR/SIRs for subcategories of non-household contact, including workplace, marketplace, and transport associated contacts. There appear to be enough data for such sub analyses.
	4. Authors estimate SAR disaggregated by symptom status and "congregation exposure of the primary case." What hypotheses are the authors evaluating in this sub analysis? What is the theoretical mechanism that would link the "congregation" as a source of COVID infection with differential risk of forward transmission? Authors should describe and reference the hypothesis specifically in the introduction as it is a focus of the paper. Authors might consider and discuss how such a finding might relate to differential responses to case or contact identification efforts.
	5. Authors could note that the estimates for the SAR for non-household contacts are in the range of available published estimates including in settings without lockdowns. The published meta-analyses of the SAR by Koh, Madewell and Lei provide

| | references to studies of transmission risks among community contacts.

6. On page 9, authors write: "We prioritized the testing of household contacts of symptomatic primary cases." What does such prioritization mean in practice? Did the prioritization occur during the period of data collection for the current analysis or afterwards? If so, findings would be biased. Is this what the authors
7. I suggest that the authors stratify the main analysis by jurisdictions, potentially grouping jurisdictions with higher / lower numbers of identified contacts per case and grouping jurisdictions with more / less rigorous protocols and consistent practices. This might test whether contact tracing capacity is significantly biasing results. |

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:

Reviewer Number	Comments	Response	Action Taken line
1	The paper adequately address the research question and can be accepted in current form.	Thank You	

Reviewer 2

Sl. No	Comments	Response	Action Taken line
Major Comments			
1	Quite a few analyses of SARS-CoV-2 secondary attack rates have previously been published, including a recent systematic review and meta-analysis. In the discussion section, the authors should place their findings within the larger body of research and highlight what additional value is provided by their analysis	We have added the previous publication on the SAR and incorporated the points in the discussion section on the evidences of systematic reviews	Line 273

2	I would suggest that the authors do not present collapsed results for the community and household contacts and explore whether these results are subject to effect measure modification. Given that some districts limited contact tracing to household contacts only, and that testing was prioritized for household contacts of symptomatic primary cases there is evidence that a head-to-head comparison such as in the risk factors section may be inappropriate. I strongly recommend reconsidering the analyses as presented in Tables 4 and 5 where the reference category is community contacts of non-congregation primary cases. The descriptive results showed that ¾ of the available contacts were community contacts, although very few became cases. The authors further state in the discussion that there was limited interaction at the community level, which also supports the need to present separately	We agree with your comment. In Table 3, we examined the SAR separately among the household and community contacts. Also, based on your suggestions, we included the confidence interval, which stated that the estimate is better. As you can observe, the SAR estimate suggested increased risk of infection across various categories. To understand the cumulative effect of each category and its dose response, we stratified into four categories (Table 4). We further stratified by symptoms status, to address the issue of effect modifier in table 5.	
3	Who were the asymptomatic primary cases? Why were they tested/how were they identified if primary cases were defined as having no established contact history?	1. All the congregation attendees were tested irrespective of symptom status, as there was lot of panic in the initial phase of the pandemic. It posed high risk of big outbreak. They were tested irrespective of the symptom status 2. Similarly, all international travelers were tested irrespective of the symptom status These persons turned positive became asymptomatic primaries.	
4	One of the main findings of this paper was that transmission was higher among contacts of primary cases exposed to the congregation. What hypotheses do the authors propose to explain why the secondary attack rate was higher in these contacts? A quarter of the primary cases were linked to this congregation which occurred early in the local epidemic (March 21-23). Some clarification is needed around the timing of the congregation exposure and the	1. At the time when this investigation was undertaken, there was no widespread community transmission. Most of the infection were among international travellers and health care workers, and they did not mingle with the community due to restrictions. Unlike international travellers, congregation participants travelled with local people. After attending the congregation, all resumed their routine work and social activities	

	implementation of strict lockdown measures. Did the earlier (congregation) cases occur during a period of less strict public health measures? Did the median number of household contacts change over the study period?	after arrival. Therefore, they are more likely to transmit to the community 2. Yes. Congregation occurred before the lockdown 3. There is no change in median number of household contacts during the study period	
Minor Comments			
1	Abstract - please clarify what you mean by congregation exposure as it only becomes clear when the body of the article is read.	We have added clarification about the congregation setting in the abstract.	Line 34
2	Strengths bullets - bullet two states that the "Majority of the contacts were tested with RT-PCR, therefore the estimates of secondary attack rate were reliable". Consider rewording; the SAR estimates were based on other assumptions that might make these estimates less reliable	All were tested for RT-PCR. The strengths bullet is updated accordingly	Line 60
3	Methods		
3a	Study setting and the COVID-19 context – please clarify what you mean by the nine administrative districts reporting maximum cases during the study period. Does this mean the analysis is focused on the nine districts with the highest case count?	The study was done in nine of the 38 administrative districts of Tamil Nadu. These nine districts reported higher number of cases during the study period. The analysis is based on the information from these nine districts	
3b	When were contacts tested relative to their quarantine period? e.g. at day 5 after last exposure to a case. Please present the median and IQR in the results	We abstracted the information from the records of the district surveillance units. The information on the date of exposure and the date of sample taken are not available for all the contacts and hence median and IQR was not calculated. This is one of the limitations of the study But as per the guidelines, all the identified contacts were tested	
3c	How was date of the contact's last exposure to the case determined? Please present the median and IQR and/or minimum and maximum periods between last exposure to the primary case and onset of	Data was extracted from the surveillance unit ad not collected from the individuals. Given that this investigation happened in the early phase, this data were not collected	

	symptoms/date of diagnosis for the secondary cases		
3d	Sampling and sample size – please clarify the study period. The study period in the abstract was March 14 to May 5, in this methods section was March 1 to May 30, and tables and title say April -May.	The data collection was done from March 14 to May 30. The data reference period was from March 1 to May 30 2020. Necessary corrections were made in the title, abstracts, methods, tables, and figures.	Lines 2, 33, 138, 367, 382, 386, 393.402
3e	Operation definitions - More details are required for these definitions		
3e – i	Primary case – does this group represent confirmed cases not named as contacts and without a relevant community exposure? Did you use symptom onset date (or date of testing for asymptomatic individuals) when determining who was the primary case in a household? How did you deal with co-primary cases (i.e. household contacts with symptom onset date within one day of primary case)?	Date of testing was taken to determine the primary case. In the early phase of the pandemic, as there was no major community transmission, the co-primary cases was not reported.	Line 159
3e – ii	Contact – please clarify definition of "exposed", e.g. unprotected contact within 2 metres for 15 or more minutes with a confirmed case	The definition for contact is given general, irrespective of high or low risk. For high-risk contacts, there is an unprotected contact of 15 min or more within 2 metres distance. For low-risk contacts, the contact may be in the same environment but not having high risk exposure.	Line 166
3e – iii	Household contact – how did you determine individuals within a household? How did you determine that contact occurred and date of last exposure to case?	The house hold contacts are individuals who are part of the family and living in the same household sharing the same kitchen. The household contact was listed based on the information collected from the primary case. And the last exposure with the case is ascertained by the health department.	Line 172
3e – iv	Cluster – what would be considered an unusual aggregation in this setting?	It refers to reporting of two or more cases. Necessary changes are made in the article	Line 161

3e - v	Secondary attack rate – please clarify this definition; I was unsure what you meant by minus the primary cases of the contacts	We agree with the comment. The definition of Secondary Attack rate is updated.	Line 182
4	Results		
4a	Description of cases and contacts – I don't understand the sentence about duration of data abstraction and am unsure what value these data add to the manuscript; suggest removal	The duration of data abstraction varies across the districts included in the study (Table 1). This is because the reporting of the first case varies across the districts and the data collection was done in the same period across these districts and hence the duration of the data collection period varies.	
4b	Secondary attack rates - please include confidence intervals in all tables	Included 95% CI values for the secondary attack rates in Tables 1, 2 & 3.	Line 376, 381 & 385
4c	Possible typo – last sentence states, "of the 598 contacts who tested positive"; should this be 599?	Error corrected and now read as '599'. The relevant table is also updated after revisiting the data.	Line 236 & 382
4d	Previous research has indicated that SARS-CoV-2 tends towards over-dispersion. How many of your primary cases did not transmit within their households?	Of the 931 primary cases, 11% (n=102) did not have household contacts.	Line 215
4e	Table 1. I would suggest that the ecological columns of population density and family size could be deleted in this table as well as the days since first case column. I would be more interested to see the contacts columns split into household and community columns	The ecological information including population density and family size argue the potential for spread of the novel disease at the time of investigation. For split of household and the community contacts, detailed analysis is made in Tables 2,3,4 & 5.	
4f	Table 2 – as with Table 1, I would recommend presenting columns for both the community and household contact types. Please include confidence intervals.	The information of household contacts and the community contacts is given for the congregation exposure and symptomatic status of the primary in tables 3, 4 & 5. Confidence interval is added in Table 3.	
4g	Did SAR vary by district?	Yes. The SAR varies from 2 to 29, with an overall SAR is 4. The district level SAR is added to the Table 1 based on the comments. The variation is due to the no. of community contacts who	Line 376

Sl.No		
	could be traced in the pandemic movement restriction situation. Although we have presented the data by districts, we have pooled the data and analysed for the key indicators.	
4h	Tables 4 and 5, please see major comment 2	Reviewed and addressed in major comment 2	

Reviewer 3:

Sl.No	Comments	Response	Action Taken line
Major			
1	What PCR target was used - single gene/multiple - which genes?	Two or more target genes (E/ RdRp/ORF1ab/ N/ S) were looked into using multiplex PCR kits.	Line 148
2	What time were individuals visited, were they tested only once? This could significantly impact on the accuracy of PCR testing of contacts - i.e missing people if tested too early or too late which could result in biased estimates. More information on how the contact tracing and testing was performed is therefore needed	1. All the congregation attendees were tested after they were traced. By the time they contacted, they have completed more than 5 to 7 days after the congregation attendance. 2. For the international travelers, they were tested if symptomatic or as per the updated guidelines 3. For others, testing was done when symptomatic or as per the guidelines 4. If any individual was tested positive, all the contacts were traced. Symptomatic contacts were tested immediately and asymptomatic were tested 5 days after the exposure 5. If the contacts were tested negative for COVID, they were observed for onset of symptoms and re-tested when symptoms appear. By implementing the above protocol, it is ensured that all the potential cases were identified	Line 108
3	Do you have data on approximate viral load (either by quantification or through the CT value of the PCR). Several previous papers have shown a	We could not abstract the information on the CT Values of all the persons included in the study, as the information was not compiled at the	

		district level. We agree with the reviewer that there is a relationship between the viral load and the risk of transmission of diseases.	
relationship between viral load and risk of onward transmission and so it would be interesting/useful to know if similar relationships were seen here. See for example a) https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30985-3/fulltext b)http://modmedmicro.nsms.ox.ac.uk/wp-content/uploads/2021/01/infectivity_manuscript_20210119_merged.pdf			
4	How was symptomatic disease defined - please include this in the methods	Symptomatic disease is defined as any COVID-19 tested positive person with H/O Fever, cough, Sore throat or breathlessness from 5 days before the date of testing	
5	Operationally how was a cluster defined?	Definition of a cluster is updated in the manuscript	Line 178
6	Operationally how was a congregation exposure defined?	Congregation exposure is defined as any person wo have attended the religious congregation event held during February and March 2020. This is included in the methods section	Line 180
Minor			
1	Page 9 - dont mix Harvard and vancouver style referencing (line 50)	Revised the reference to Vancouver (Ref 18)	

Reviewer 4:

Sl. No	Comments	Response	
1	There are multiple references and analyses of risks in relation to "congregation exposure;" however, "congregation exposure" term is never explicitly and operationally defined in the methods. There is an oblique discussion in the introduction to the paper about a cluster associated with a religious congregation in New Delhi that began on February 9, 2021. Presumably, one assumes that "congregation exposure" refers to this particular religious congregation;	We agree with your comments. The definition of the congregation exposure is given in the methods section.	Line 180

	however, the term congregation exposure could many have other meanings (i.e., any gatherings) and should be more precisely defined. The source of information for 'congregation exposure in the primary case" should also be precisely described given its import in the article		
2	The data for this analysis comes from government public health case identification and contact tracing efforts. Authors should summarize written protocols for case identification, contact identification, and contact follow up used by these government agencies and ideally provide as supplementary materials. If the protocols varied by jurisdiction, their salient aspects could be synthesized in a table organized by jurisdiction. Salient information in the protocols includes the criteria for contact (e.g., time relative to the case identification data, duration of contact, proximity), period and duration of contact follow-up following contact identification, and timing of symptom and biological disease assessments relative to the timing of contact identification	The data is collected from the government public health department. There is a uniform guideline for the entire state of Tamil Nadu, developed and disseminated by Ministry of Health, Govt of India. Testing Strategy: Testing Strategy (icmr.gov.in) Protocols for case identification, contact reacing and follow up: Standard Operating Procedures SOPs :: National Centre for Disease Control (NCDC)	
3	Most "contacts" were non-household. It would be very useful given the paucity of scholarship on exposure risks among non-household community and workplace contacts to estimate SAR/SIRs for subcategories of non-household contact, including workplace, marketplace, and transport associated contacts. There appear to be enough data for such sub analyses.	The information sub-category of the contacts is not available and hence could not be analyses. This is a data limitation in the analysis.	
4	Authors estimate SAR disaggregated by symptom status and "congregation exposure of the primary case." What hypotheses are the authors evaluating in this sub analysis? What is the theoretical mechanism that would link the "congregation" as a source of COVID infection with differential risk of forward transmission? Authors should describe and reference the hypothesis specifically in the introduction as it is a	At the time this investigation was undertaken, there was no wide spread community transmission. Most of the infection were among travelers and health care workers, and they did not mingle with the community due to restrictions. Unlike international travelers, congregation participants travelled with other local travelers and after attending the congregation, all	Line 121 & 309

	focus of the paper. Authors might consider and discuss how such a finding might relate to differential responses to case or contact identification efforts.	resumed social and work-related activities after arrival. Therefore, they are more likely to transmit to the community	
5	Authors could note that the estimates for the SAR for non-household contacts are in the range of available published estimates including in settings without lockdowns. The published meta-analyses of the SAR by Koh, Madewell and Lei provide references to studies of transmission risks among community contacts.	We agree with your comments. We have added the articles and discussed the same	Line 273
6	On page 9, authors write: "We prioritized the testing of household contacts of symptomatic primary cases." What does such prioritization mean in practice? Did the prioritization occur during the period of data collection for the current analysis or afterwards? If so, findings would be biased. Is this what the authors	During the study period, all contacts were tested irrespective of the symptoms to identify the cluster of cases. However, as the COVID-19 epidemic progressed, the testing capacity could not increase at the pace at which the cases increased. Therefore, it was recommended for prioritization of symptomatic contacts to reduce the spread. And this was consistent with the evidence that came out of our study	Line 328
7	I suggest that the authors stratify the main analysis by jurisdictions, potentially grouping jurisdictions with higher / lower numbers of identified contacts per case and grouping jurisdictions with more / less rigorous protocols and consistent practices. This might test whether contact tracing capacity is significantly biasing results.	The guidelines for contact tracing and isolation were similar for all the districts. There was uniformity in the protocol. However, the number of contacts identified by case are influenced by multiple factors such as cooperation of the community, trained work force and other logistic issues. And this cluster occurred in early phase of pandemic. So, all the protocols and the covid control strategies are equally new to all the districts. Therefore, the authors feel that the stratification is not relevant.	

VERSION 2 – REVIEW

REVIEWER	Marks, Michael London School of Hygiene and Tropical Medicine, Clinical Research Department
REVIEW RETURNED	18-Sep-2021

GENERAL COMMENTS	I am satisfied they have responded to my earlier comments