PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Seroprevalence of SARS-CoV-2 specific IgG antibodies in Kashmir, India, seven months after the first reported local COVID-19 case: results of a population-based seroprevalence survey from October-November, 2020
AUTHORS	Khan, S Muhammad; Qurieshi, Mariya; Haq, Inaamul; Majid, Sabhiya; Ahmad, Javid; Ayub, Taha; Bhat, Ashfaq; Fazili, Anjum; Ganai, Abdul; Jan, Yasmeen; Kaul, Rauf-ur-Rashid; Khan, Zahid; Masoodi, Muneer; Mushtaq, Beenish; Nazir, Fouzia; Nazir, Muzamil; Raja, Malik; Rasool, Mahbooba; Asma, Anjum; Ayoub, Shifana; Aziz, Munazza; Bhat, Arif; Chowdri, Iqra; Ismail, Shaista; Kawoosa, Misbah; Khan, Mehvish; Khan, Mosin; Kousar, Rafiya; Lone, Ab; Nabi, Shahroz; Obaid, Mohammad; Qazi, Tanzeela; Sabah, Iram; Sumji, Ishtiyaq

VERSION 1 – REVIEW

REVIEWER	Stoto, Michael Georgetown University
REVIEW RETURNED	12-Jun-2021

GENERAL COMMENTS	This is an interesting paper that uses valid and appropriate epidemiological and statistical methods to address an important and timely topic. It could be a useful addition to the growing literature of seroprevalence studies, some of which is cited references #5 – #15. The analysis is, however, seriously out of date. I am not sure about Kashmir particularly, but in all of India the total number of cases is more than three times higher now than at the time the survey was in the field. Thus, the overall seroprevalence estimate is primarily of historical interest. What's interesting about the results, and may be of more enduring value, is the analysis of the relationship among being seropositive, having had COVID-19 symptoms, having been reported as a case, having been tested, and the result of that test. These relationships are presented in the tables and discussed in the text (p. 8, ll. 17-33 and p. 10, ll. 10-21) mostly in a pairwise fashion, which makes is difficult to fully understand. It would be helpful to lay out the relationship among them in graphical terms such as in Angulo et al. (JAMA Network Open 2021 4(1):e2033706) or Holtgrave et al. (Ann Epidemiol. 2020 48:9-14. doi: 10.1016/j.annepidem.2020.06.010). Other points The infection fatality rate (IFR) of 342·1 per million infections seems very low to me. This rate is more commonly represented as a percentage, so this estimate is 0.034%. In a systematic review of 61 studies, Ioannidis (Bull World Health Organ 2021 99:19–33F. doi:

	http://dx.doi.org/10.2471/BLT.20.265892) finds that IFR ranged from 0.00% to 1.63%, with a median of 0.27%. Two studies from India had IFRs of 0.06% and 0.09% in July, 2020. So, references #24 and #31 notwithstanding, this should be looked at more carefully. In addition to not having quantified the test validity in-house (p. 10, ll. 28-30), the analysis also did not take into account the precision of the estimates of sensitivity and specificity. This can be done in a number of ways. Rosenberg et al. (Annals of Epidemiology (2020), doi: 10.1016/j.annepidem.2020.06.004) do it using a sensitivity analysis. Meyer et al. have developed a formal Bayesian analysis (https://medrxiv.org/cgi/content/short/2021.03.04.21252939v1)

REVIEWER	Roederer, Thomas Epicentre, Epidemiology
REVIEW RETURNED	20-Jun-2021

GENERAL COMMENTS	The authors of this study, and of this manuscript, have done a remarkable job. The topic is not well described in the literature, and if it has been, it is by the exact same team. The methods are clearly presented, even if they may be lacking in detail (but this is probably due to the word limit). The results are clear and speak for themselves (the tables need to be redone to be cleared). The main flaw of this article lies in the discussion: surprisingly the authors do not really put the results in perspective with their previous work. The limitations are barely touched upon, especially the fact of insisting so much on the total number of cases or deaths in the region, when we know well the reporting problems India has been experiencing since the beginning of the crisis. Finally, but this is my humble opinion, this article feels lackluster and too late, since it describes a situation dating back to the end of 2020, while the current situation of the COVID-19 epidemic in India is totally different (out of control), especially because of new variants. These aspects are not discussed at all, as is the matter of vaccination. In the current state of the manuscript, I think the authors can improve the points discussed above (and detailed further in the attachment), which should not cause too much work.

VERSION 1 – AUTHOR RESPONSE

Thank you for the opportunity to submit a revision of our manuscript to BMJ Open. We appreciate your efforts in providing valuable feedback on our manuscript. We are grateful to the reviewers and appreciate the time and effort they have put in to provide their insightful comments. Based on the suggestions provided by you and the reviewers we have made several changes to the manuscript. Here is a point-by-point response to reviewers' comments and concerns.

Reviewer: 1 Prof. Michael Stoto, Georgetown University			
	What's interesting about the results, and may be of more enduring value, is the analysis of the	We are really thankful for this comment and have added one figure (Figure 3) in this context. Hope the	Figure 3 added.

	relationship among being seropositive, having had COVID-19 symptoms, having been reported as a case, having been tested, and the result of that test. These relationships are presented in the tables and discussed in the text (p. 8, ll. 17-33 and p. 10, ll. 10-21) mostly in a pairwise fashion, which makes is difficult to fully understand. It would be helpful to lay out the relationship among them in graphical terms such as in Angulo et al. (JAMA Network Open 2021 4(1):e2033706) or Holtgrave et al. (Ann Epidemiol. 2020 48:9-14. doi: 10.1016/j.annepidem.2020.06.010).	figure makes things easier to read.	
	The infection fatality rate (IFR) of 342·1 per million infections seems very low to me. This rate is more commonly represented as a percentage, so this estimate is 0.034%. In a systematic review of 61 studies, Ioannidis (Bull World Health Organ 2021 99:19–33F. doi: http://dx.doi.org/10.2471/BLT.20.265892) finds that IFR ranged from 0.00% to 1.63%, with a median of 0.27%. Two studies from India had IFRs of 0.06% and 0.09% in July, 2020. So, references #24 and #31 notwithstanding, this should be looked at more carefully.	The Infection Fatality Rate (IFR) now reads as percentage, instead of per million, throughout the article. Thanks for the suggestion, since now the findings of our study will be more easily comparable across studies. We have rewritten the paragraph on IFR in the discussion section in light of the references provided.	The paragraph number 10 in the discussion section now reads, "We estimated an infection fatality rate of 0.034% (95% CI 0.032 – 0.037). The infection fatality rate in SARS-CoV-2 infection has been reported to range from as low as 0.00% to 1.63%. Our estimates of the infection fatality rate are low as compared to estimates from several Indian studies. Under-reporting of COVID-19 deaths because of non-uniform definition for a 'COVID-19 death' may falsely lower the infection fatality rates. The infection fatality rate is, however, known to be lower in

			developing nations. In developed nations like the United States and many European countries, a higher infection fatality rate has been reported."
	In addition to not having quantified the test validity in-house (p. 10, ll. 28-30), the analysis also did not take into account the precision of the estimates of sensitivity and specificity. This can be done in a number of ways. Rosenberg et al. (Annals of Epidemiology (2020), doi: 10.1016/j.annepidem.2020.06.004) do it using a sensitivity analysis. Meyer et al. have developed a formal Bayesian analysis (https://medrxiv.org/cgi/content/short/2021.03.04.21252939v1)	Although we did not perform a Bayesian analysis, we now report the sensitivity analysis. We have added Table 3 in the manuscript in this context.	Table 3 added in the manuscript.
Reviewer: 2 Dr. Thomas Roederer, Epicentre	The authors of this study, and of this manuscript, have done a remarkable job. The topic is not well described in the literature, and if it has been, it is by the exact same team. The methods are clearly presented, even if they may be lacking in detail (but this is probably due to the word limit). The results are clear and speak for themselves (the tables need to be redone to be cleared). The main flaw of this article lies in the discussion: surprisingly the authors do not really put the results in perspective with their	Thanks for the initial comment. Your comment are valuable and have helped to make the manuscript a lot better, hopefully. We have re-written the discussion part of the manuscript and added few more limitations.	

	previous work. The limitations are barely touched upon, especially the fact of insisting so much on the total number of cases or deaths in the region, when we know well the reporting problems India has been experiencing since the beginning of the crisis.		
	Finally, but this is my humble opinion, this article feels lacklustre and too late, since it describes a situation dating back to the end of 2020, while the current situation of the COVID-19 epidemic in India is totally different (out of control), especially because of new variants. These aspects are not discussed at all, as is the matter of vaccination. In the current state of the manuscript, I think the authors can improve the points discussed above (and detailed further below), which should not cause too much work.	COVID-19 has been changing worldwide with peaks appearing very now and then. As we speak today, the so called “second wave” has receded and we are awaiting the third wave. We have tried to improve the article in light of the present changes and the valuable comments.	Statements on virus variants, and vaccination have been added at several places in the manuscript. We have improved the article in light of the comments by all reviewers.
	Background We don’t really know the situation of Kashmir stat in November 2020 at the time of the survey. Was this during a surge of the outbreak? in between?	Thanks for this comment. At the time of writing the first manuscript draft we were unsure whether to include such information or skip it. Thanks to your comment, we have added one figure (Figure 4) which details the reported cases and deaths in Kashmir since the start of the pandemic. Hope this gives more perspective to the article.	Figure 4 has been added in the manuscript.
	Methods *Objectives are not really stated: seroprevalence, ok,	Estimating the seroprevalence was the primary objective of the study. Because	The ‘objectives’ paragraph at the end of the “introduction”

	but why? Are there no other aim to this study ?	of word limits we included only the primary objective in the manuscript draft. We have reworded the 'objectives' statement. Our secondary objectives were 1) to assess the relationship between various demographic variables and seroprevalence; 2) To estimate the number of infections per reported case; and 3) To estimate the IFR.	section now reads "We designed this survey with the primary objective to estimate the seroprevalence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) specific IgG antibodies in the adult population of Kashmir valley."
	Sample size calculation looks ok. I would have liked to see the sample size per district somewhere though. Where is the justification of the 20% hypothesized for seroprevalence ?	We have re-written the sample size paragraph. The 20% seroprevalence was speculated based on a previous survey in the central of Srinagar. We have added the requisite reference and re-written the sample size paragraph following comments by the editor as well.	The sample size paragraph now reads "Based on the results of a previous study conducted in July 2020, we speculated that, by October 2020, the prevalence would have increased to around 20%.[16] We calculated the minimum sample size based on an anticipated seroprevalence of 20%, an absolute precision of 2%, and a design effect of 2. We used OpenEpi to make sample size calculations.[17] We adjusted the sample size for a possible non-response of 10% to obtain a minimum size of 3376. We decided to select 3600 individuals from nine of the ten districts (except district Srinagar). To obtain precise estimates for district Srinagar, sample size estimation was made for the district separately. We used a design effect of 1.5, an anticipated seroprevalence of 20%, and absolute precision of 2% to obtain a sample size of 2302 for the district,

			which was further increased to 2400 to account for non-response. We thus targeted a total sample size of 6000 (3600 + 2400)."
	How were the locations in each cluster randomly selected ? Spatial sampling ? PPS as well ? This is not very clear. Describe more the process to include household, was it systematic ? by proximity ? This whole section needs clearer details.	From our experience of previous surveys were we faced much difficulty in spatial sampling, we chose a different and feasible approach to to select households within a randomly chosen cluster. Clusters were chosen using PPS within each stratum. Each cluster was then divided into four almost equal areas and a central location within each of the four areas was identified. We instructed the field team to choose a random direction and the first household in that randomly chosen direction. Thereafter, households were chosen consecutively till the requisite sample size was achieved.	We have made this change in the 'participants' sub-section of the 'methods' section - "We divided each selected cluster into four equal areas and chose a central location within each of the four areas as the starting point. Thereafter, we approached consecutive households to enroll at least ten eligible participants."
	I understand the aim of inferring Infection Fatality Ratio and case loads, but I feel like it should have been made through proper modelling (Bayesian methods). Considering the under-reporting of cases in India, I supposed reporting of cases and deaths in Kashmir is also problematic.	We have re-written the manuscript sections on IFR and mention has been made in the limitations section as well. We express our inability to perform the Bayesian modelling. Further, we believe the under-reporting in Kashmir, especially deaths, is not as problematic as in some other Indian states.	The IFR section has been re-written in the methods section and the limitations sub-section of the discussion.
	Why not Clopper-Pearson method for computing Confidence Interval ? I acknowledge I am not very aware of the Agresti-Coull, but where is the justification for this choice ?	The Agresti-Coull interval is less conservative in comparison to Clopper-Pearson and has been recommender for large samples. [Please refer Five Confidence Intervals for	We have added a reference to the choice. Reference number

		Proportions That You Should Know About | by Dr. Dennis Robert MBBS, MMST | Towards Data Science for details; and the references at the end of the article for more discussion]. Statist. Sci. 16(2): 101-133 (May 2001). DOI: 10.1214/ss/1009213286 is a particularly interesting read.	
	Table 2 is not well formatted and quite hard to read. I suggest to redo it.	Table 2 has been re-formatted. Hopefully, it'll be more readable now.	
	Women were more exposed than men, this could look weird but it happens, I expect an attempt at explanation in the discussion.	The seroprevalence in females was not very different from males, and it was not statistically significantly different as well. Its just a chance occurrence, so we did not feel the need to discuss it.	
	Seroprevalence increases with age : here again, this is not usual. Does it have to do with poor adherence to safety measures ? The discussion should address this point.	We have discussed the association of age and seroprevalence. Unfortunately, we did not measure the adherence to prevention measures in our study sample.	
	Roughly 36% of seropositives never reported symptoms, didn't know anyone with COVID or got tested by PCR. Are they always the same 36%? I would have like some cross tables here. The authors definitely could dig further in those data.	Honestly, we fail to understand this comment. But since it points to something amiss with the way wrote things in the original draft, and because of comments by another reviewer as well, we have added a figure (Figure 3) to make things more clear and understandable.	Figure 3 has been added in the manuscript.
	While I understand the intention with computing IFR or CFR, I have a hard time trusting the reported number of cases or deaths.	We have mentioned the under-reporting of deaths in the limitations section. IFR is a concern and should be read with caution.	We have addd this statement to the limitations section, "Lack of reliable death counts is another potential limitation. This may have led to an underestimation of the infection fatality

			rate. We did not perform any adjustment for death counts.”
	Where are the sensitivity analyses : seroprevalence using different sensitivity/specificity for the Abbott assay. Maybe different inferences on CFR/IFR using more realistic data from the literature as well ?	We have added table3 which reports the sensitivity analyses.	Table 3 has been added.
	Discussion The authors do put their results into perspective of similar studies published in the literature. But they miss important points : I would have expected comparison of the seroprevalence results with other parts of India, or with the other surveys published by the authors from the same team : seroprev in health workers 3.6%, Ref 56): https://doi.org/10.1371/journal.pone.0239303 Or this one, IJID paper on seroprevalence in india, with some coauthors in common, 3e d’une série de 3 serosurveys -> dec-jan seroprev in healthcare workers was 25.6% Or even this one in the same area by the same team ! https://dx.doi.org/10.18203/2394-6040.ijcmph20211236 This paper could also be discussed or at least be mentioned : Lancet GH paper (nationwide 6.6%) : https://doi.org/10.1016/S2214-109X(20)30544-1	We have added one new paragraph in the discussion section addressing this issue.	Following paragraph has been added in the discussion section (Discussion para 4) “Comparison with previous reports suggests that the seroprevalence has increased almost ten-fold since July 2020.[16,27] The second of the three nationwide seroprevalence surveys in India conducted in August-September 2020 reports an overall seroprevalence of 6.6% ranging from 5.2% in rural areas to 16.9% in urban slums.[28] A nationwide survey conducted in December 2020-January 2021 reported an overall seroprevalence of 24.1% ranging from 4.9% - 44.4% across districts.[29] Kashmir is thus not a low-infection area. Being an oft-visited tourist area, Kashmir is at an increased risk of infection transmission. Adherence to COVID appropriate behavior (use of face masks in public, frequent handwashing, physical and social distancing) has been poor. With

			the introduction of the COVID-19 vaccination program in January 2021 and the emergence of a 'second wave' in Kashmir in April 2021, the seroprevalence estimates are expected to increase in the future."
	The authors never address the issue of under-reporting of cases in India (in general), is it different in Kashmir ? Sources for this issue : https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)31857-2/fulltext https://www.nytimes.com/interactive/2021/05/25/world/asia/india-covid-death-estimates.html https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7818846/ from 20 to 300M cases, and more than double the deaths ! Conclusions about CFR and IFR are therefore to take really cautiously.	We have added a statement to this effect in the limitations section.	Following statements have been added in the limitations part; "Lack of reliable death counts is another potential limitation. This may have led to an underestimation of the infection fatality rate. We did not perform any adjustment for death counts."
	Overall: Statistical Analyses are a bit too basic : they could definitely have done more, maybe some logistic/Poisson regression to dig into risk factors much more.	Exploring the risk factors and estimating their effect on seroprevalence was not a primary objective of the study. We did perform a univariable analysis to give an overview about the possible risk factors of seropositivity.	
	I would have liked a map, since I don't know they area very well. I would have shown the seroprevalences by district (and confidence intervals) on a map as well.	Changes as rightly suggested have been made. We have added a map of the selected area, have shown its location within India, and have also shown the seroprevalence and its 95% CI for each district.	Figure 1 has been added.

	I really congratulate the authors for the quality work: the topic is interesting, the article is well written and overall methods are sound. Unfortunately, this feels lacklustre, knowing the current situation in India, this paper feels late. The authors didn't discuss variants, vaccination and current situation. Maybe they could add a few sentences ?	We have added few sentences about vaccination and virus variants in the discussion part of the manuscript.	Statements added: Discussion para 3 "The emergence of several Variants of Concern and the introduction of COVID-19 vaccination will also influence population immunity."

VERSION 2 – REVIEW

REVIEWER	Stoto, Michael Georgetown University
REVIEW RETURNED	17-Jul-2021

GENERAL COMMENTS	As I noted in my earlier review, this is an interesting paper that uses valid and appropriate epidemiological and statistical methods to address an important and timely topic. It could be a useful addition to the growing literature of seroprevalence studies, some of which is cited. However, despite the revisions, I continue to believe that the analysis is seriously out of date. I am not sure about Kashmir particularly, but in all of India the total number of cases is more than three times higher now than at the time the survey was in the field. Thus, the overall seroprevalence estimate is primarily of historical interest. The authors did not respond to a similar comment in my original review, but did to a similar comment from the other reviewer, and adding information from other studies in India earlier and later helped address issue. But as had become abundantly clear the major part of India's COVID-19 cases to date have been in the current year. Consequently, the conclusion that "A large proportion of the population remains susceptible to the infection" (Abstract) may not be true in July, 2021. The authors are correct that "The experience of a second wave of COVID-19 in April-June 2021, the appearance of virus variants, and the introduction of vaccination programs warrant robust surveillance of the epidemic," but that is hardly the conclusion of the seroprevalence study described in this paper. I appreciate the addition of Figure 3, however, I do not believe that it goes far enough to describe the complex relationship between infections, testing, reported cases, and reported deaths. The table, figure, and text do little to clarify this relationship. I still believe that the graphical analyses in Angulo et al. (JAMA Network Open 2021 4(1):e2033706) and Holtgrave et al. (Ann Epidemiol. 2020 48:9-14. doi: 10.1016/j.annepidem.2020.06.010) should be used as a model.

	The discussion about the low infection fatality rate (IFR) is related to this issue. I do not believe that simply adding a statement that "The infection fatality rate is, however, known to be lower in developing nations" (p. 11, ll. 36-37) is sufficient. The second reviewer raised the same issue in noting that there is likely underreporting of cases and deaths, and the authors responded by adding some text in the Limitations section. It would be more appropriate to connect this to the discussion of the low IFR a few lines above. Isn't it possible, or even likely, that other developing nations have low IFR estimates because they too have incomplete mortality estimates? Finally, I appreciate the addition of the correction for sensitivity and specificity, but I do not believe that the authors have implemented even the standard method correctly. Referring to Table 3, there should be four columns in the sensitivity analysis, corresponding to the upper and lower confidence intervals of the assumed sensitivity and specificity. Consequently, the range cited in the text (p. 6, l. 24) is not correct. But then the authors go on to discuss the results as if this analysis had never been done.

REVIEWER	Roederer, Thomas Epicentre, Epidemiology
REVIEW RETURNED	03-Aug-2021

GENERAL COMMENTS	The authors have addressed most of my concerns and questions. The article now feels much smoother to read, limitations are now discussed, tables are much clearer and I do appreciate the map, the sensitivity analysis and the additional details on methods. There must still be improvements on the english to do, but as it is, I recommend the article for acceptance.

VERSION 2 – AUTHOR RESPONSE

Thank you for the opportunity to submit a revision of our manuscript to BMJ Open. We appreciate your efforts in providing valuable feedback on our manuscript. We are grateful to the reviewers and appreciate the time and effort they have put in to provide their insightful comments. Based on the suggestions provided by the reviewers we have made several changes to the manuscript. Here is a point-by-point response to reviewers' comments and concerns.

Comments from	Comment	Response	Changes made in manuscript
Reviewer: 1 Prof. Michael Stoto, Georgetown University	However, despite the revisions, I continue to believe that the analysis is seriously out of date. I am not sure about Kashmir particularly, but in all of India the total number of cases is more than three times higher now than at the time the survey was in the field. Thus, the	The survey was conducted in October-November 2020 and reflects the seroprevalence estimates towards the end of the "first wave" of the epidemic in Kashmir. The results of the survey reflect the population seroprevalence almost 7 months after the appearance of the first local case of COVID-	

	overall seroprevalence estimate is primarily of historical interest. The authors did not respond to a similar comment in my original review, but did to a similar comment from the other reviewer, and adding information from other studies in India earlier and later helped address issue. But as had become abundantly clear the major part of India's COVID-19 cases to date have been in the current year.	19. Though the analysis seems out of date, we believe that our results provide significant data and information which shall be of use in future infectious disease pandemics/epidemics. With the ever-increasing availability and use of big data analysis techniques, our work is a small but significant addition to possible future analyses.	
	Consequently, the conclusion that "A large proportion of the population remains susceptible to the infection" (Abstract) may not be true in July, 2021.	As of 13-08-2021, only 8.7% of the Indian population is fully vaccinated against COVID-19 [www.covid19india.org]. Even after the "second wave" of COVID-19 in Kashmir, there is a significant proportion of people who are not immune to the infection. Infection-induced immunity, in COVID-19, is known to wane off over time. Introduction of Variants of Concern and Variants of High Consequence may shift the balance again in the future.	

		Therefore, we believe, this particular statement "A large proportion of the population remains susceptible to the infection" is still relevant.	
	The authors are correct that "The experience of a second wave of COVID-19 in April-June 2021, the appearance of virus variants, and the introduction of vaccination programs warrant robust surveillance of the epidemic," but that is hardly the conclusion of the seroprevalence study described in this paper	This statement is, sort of, a recommendation. It reflects what we need to keep doing and keep doing it better.	
	I appreciate the addition of Figure 3, however, I do not believe that it goes far enough to describe the complex relationship between infections, testing, reported cases, and reported deaths. The table, figure, and text do little to clarify this relationship. I still believe that the graphical analyses in Angulo et al. (JAMA Network Open 2021 4(1):e2033706) and Holtgrave et al. (Ann Epidemiol. 2020 48:9-14. doi: 10.1016/j.annepidem.2020.06.010) should be used as a model.	The primary focus of our study was not to describe in detail the complex relationship between infections, testing, cases, and deaths, which we intend to do in a separate analysis of a more recent survey with additional data points. However, we are attaching Figure 5 which will, hopefully, serve the purpose sought.	Figure 5 added.
	The discussion about the low infection fatality rate (IFR) is	We have further revised the discussion on Infection Fatality	The paragraph on IFR in the discussion section now reads "We

	related to this issue. I do not believe that simply adding a statement that "The infection fatality rate is, however, known to be lower in developing nations" (p. 11, ll. 36-37) is sufficient. The second reviewer raised the same issue in noting that there is likely underreporting of cases and deaths, and the authors responded by adding some text in the Limitations section. It would be more appropriate to connect this to the discussion of the low IFR a few lines above. Isn't it possible, or even likely, that other developing nations have low IFR estimates because they too have incomplete mortality estimates?	Rate. We acknowledge that our IFR estimates could be lesser than the true IFR in the population. Unfortunately, at present, we do not have an informed method to make our estimates better. Estimating IFR is complex as it relies heavily on the reported number of deaths. Chris Kenyon has suggested that the IFR is influenced by COVID-19 epidemic intensity and may not be a constant number [*Kenyon C. COVID-19 Infection Fatality Rate Associated with Incidence-A Population-Level Analysis of 19 Spanish Autonomous Communities. Biology (Basel) 2020;**9**:1–4. doi:10.3390/BIOLOGY 9060128*]. IFR is strongly related to age and hence the age structure of a population will influence the overall IFR. Kashmir has a comparatively young population with 7.4% above 60 years and 3.2% above 70 years of age.	estimated an infection fatality rate of 0.034% (95% CI 0.032 – 0.037). The infection fatality rate in SARS-CoV-2 infection has been reported to range from as low as 0.00% to 1.63%.[36] Our estimates of the infection fatality rate are low as compared to estimates from several Indian studies.[5,28,37] Under-reporting of COVID-19 deaths because of non-uniform definition for a 'COVID-19 death' may falsely lower the infection fatality rates.[38] Many other factors can influence the infection fatality rate in SARS-CoV-2 infection – the quality of available health facilities, the age structure of the population, and COVID-19 epidemic intensity.[39,40] Developing countries usually have a younger population as compared to the developed countries and Kashmir is not an exception. However, because of the possibility of under-reporting of COVID-19 deaths, the true infection fatality rate in Kashmir may be higher than our estimates. The infection fatality rate is, however, known to be lower in developing nations.[30,41] In

			developed nations like the United States and many European countries, a higher infection fatality rate has been reported.[30,42]"
	Finally, I appreciate the addition of the correction for sensitivity and specificity, but I do not believe that the authors have implemented even the standard method correctly. Referring to Table 3, there should be four columns in the sensitivity analysis, corresponding to the upper and lower confidence intervals of the assumed sensitivity and specificity. Consequently, the range cited in the text (p. 6, l. 24) is not correct. But then the authors go on to discuss the results as if this analysis had never been done.	In Table 3, we report the two possible extremes for seroprevalence using the 95% CI for test sensitivity (Se) and specificity (Sp). Although four combinations of Se and Sp are possible (LLSe&LLSp, ULSe&ULSp, LLSe&ULSp, ULSe&LLSp), the seroprevalence derived from using LLSe&ULSp gives the highest extreme and the one derived from ULSe&LLSp gives the lowest extreme. The seroprevalence derived from LLSe&LLSp and ULSe&ULSp lie between the two extremes of LLSe&ULSp and ULSe&LLSp. {UL=Upper Limit, LL=Lower Limit}. Similar approach has been used by Rosenberg ES et al [*Rosenberg ES, Tesoriero JM, Rosenthal EM, et al. Cumulative incidence and diagnosis of SARS-CoV-2 infection in New York. Ann Epidemiol. 2020 Aug;48:23-29.e4. doi: 10.1016/j.annepidem.2020.06.004.*] The range of bias estimated through sensitivity analysis was fairly narrow. We thus skipped the discussion to curtail	

		the number of words in the manuscript.	
Reviewer: 2 Dr. Thomas Roederer, Epicentre	The authors have addressed most of my concerns and questions. The article now feels much smoother to read, limitations are now discussed, tables are much clearer and I do appreciate the map, the sensitivity analysis and the additional details on methods. There must still be improvements on the English to do, but as it is, I recommend the article for acceptance.	Thank you for the efforts you put into reviewing our manuscript. We deeply appreciate your comments which have improved the manuscript significantly.	

VERSION 3 – REVIEW

REVIEWER	Stoto, Michael Georgetown University
REVIEW RETURNED	22-Aug-2021

GENERAL COMMENTS	As I noted in my earlier reviews, this is an interesting paper that uses valid and appropriate epidemiological and statistical methods to address an important and timely topic. It could be a useful addition to the growing literature of seroprevalence studies, some of which is cited. However, despite both revisions, I continue to believe that the analysis is seriously out of date. I am not sure about Kashmir particularly, but in all of India the total number of cases is more than three times higher now than at the time the survey was in the field. Thus, the overall seroprevalence estimate is primarily of historical interest. Regarding the back and forth about the proportion of the Kashmir population that remains susceptible to infection in August, 2021, the key point is that the seroprevalence survey conducted almost a year earlier, before one or more major waves of infection, is not informative. Consequently, although I agree with authors that “The experience of a second wave of COVID-19 in April-June 2021, the appearance of virus variants, and the introduction of vaccination programs warrant robust surveillance of the epidemic,” this is hardly the conclusion of the seroprevalence study described in this paper.

	The addition of Figure 5 does not solve the problem I noted earlier about the need to describe the complex relationship between infections, testing, reported cases, and reported deaths. This figure is not similar to the graphical analyses in Angulo et al. (JAMA Network Open 2021 4(1):e2033706) and Holtgrave et al. (Ann Epidemiol. 2020 48:9-14. doi: 10.1016/j.annepidem.2020.06.010) which I suggested as a model. The discussion about the low infection fatality rate (IFR) resolves the issues that I had earlier raised. Finally, regarding the correction for sensitivity and specificity, I am glad to hear that the authors did implement the standard method correctly. The fact that they did this should be noted in the text, and the results reported, not just skipped because they claim that the resulting interval was "fairly narrow."

VERSION 3 – AUTHOR RESPONSE

Thank you for the opportunity to submit a revision of our manuscript to BMJ Open. We appreciate your efforts in providing valuable feedback on our manuscript. We are grateful to the reviewers and appreciate the time and effort they have put in to provide their insightful comments. Based on the suggestions provided by the reviewers, we have made several changes to the manuscript. Here is a point-by-point response to reviewers' comments and concerns.

Comments from	**Comment**	**Response**	**Changes made in the manuscript**
Reviewer: 1 Prof. Michael Stoto, Georgetown University	As I noted in my earlier reviews, this is an interesting paper that uses valid and appropriate epidemiological and statistical methods to address an important and timely topic. It could be a useful addition to the growing literature of seroprevalence studies, some of which is cited. However, despite both revisions, I continue to believe that the analysis is seriously out of date. I am not sure about Kashmir particularly, but in all of India the total number of cases is more than three times higher now than at the	We have made changes in the manuscript to clarify that the current situation in Kashmir is different from what we present through our study. We have added a paragraph at the start of the discussion section. We hope this addresses the issue.	Following text was added. ***"We report the results of a seroprevalence survey conducted in Kashmir from October-November 2020, seven months after the appearance of the first local COVID-19 case. The COVID-19 pandemic is rapidly evolving worldwide. In Kashmir, several important events happened since we completed our survey. From 16 Jan 2021, COVID-19 vaccination was introduced in a phased manner. Healthcare workers were given preference during the first phase. From***

	time the survey was in the field. Thus, the overall seroprevalence estimate is primarily of historical interest.		***01 Mar 2021, the vaccine was made available for people ≥60 years of age and those with chronic diseases in the age group of 45-59 years. However, especially during the early phases of the COVID-19 vaccination campaign, many people were hesitant to receive the vaccine doses. During the same time, SARS-CoV-2 Variants of Concern began to emerge and circulate. The daily number of COVID-19 cases started to rise again. The 'second wave' in April 2021 was more explosive than the 'first wave' at the beginning of the pandemic. The fear of the disease had diminished, and COVID appropriate behaviour was no more a norm. The government and the people were caught unawares. There were several reports of a possible 'second infection' and reports of cases among previously vaccinated individuals. Given these developments, the current seroprevalence in Kashmir will be higher than what we report in this study."***

	Regarding the back and forth about the proportion of the Kashmir population that remains susceptible to infection in August, 2021, the key point is that the seroprevalence survey conducted almost a year earlier, before one or more major waves of infection, is not informative. Consequently, although I agree with authors that "The experience of a second wave of COVID-19 in April-June 2021, the appearance of virus variants, and the introduction of vaccination programs warrant robust surveillance of the epidemic," this is hardly the conclusion of the seroprevalence study described in this paper.	We agree that the statement should not be perceived as the conclusion of our findings. Accordingly, we have removed the statement from the conclusion part in the abstract and replaced it.	The 'conlcusions' part in the abstract section now reads, ***"During the first seven months of the COVID-19 epidemic in Kashmir valley, approximately 37% of individuals were infected. The reported number of COVID-19 cases was only a small fraction of the estimated number of infections. A more efficient surveillance system with strengthened reporting of COVID-19 cases and deaths is warranted."***
	The addition of Figure 5 does not solve the problem I noted earlier about the need to describe the complex relationship between infections, testing, reported cases, and reported deaths. This figure is not similar to the graphical analyses in Angulo et al. (JAMA Network Open 2021 4(1):e2033706) and Holtgrave et al. (Ann Epidemiol. 2020 48:9-14. doi: 10.1016/j.annepidem.2	Figure 5 in our manuscript has been adapted from Holtgrave et al. In our study, we are not comparing different races or sections of the society, and we don't have data for the number of hospitalizations. Nevertheless, we believe Figure 5 provides nice graphical information about the relationship between infections, reported cases, and reported deaths. We couldn't display the relationship between infections and testing	

	020.06.010) which I suggested as a model.	in this scenario because a person gets tested more than once, so we couldn't estimate the number of persons in Kashmir who have been tested till now.	
	The discussion about the low infection fatality rate (IFR) resolves the issues that I had earlier raised.		
	Finally, regarding the correction for sensitivity and specificity, I am glad to hear that the authors did implement the standard method correctly. The fact that they did this should be noted in the text, and the results reported, not just skipped because they claim that the resulting interval was "fairly narrow."	We have made few changes. Statements about the sensitivity analyses now appear in the methods, results, and discussion section. In the results section "Upon sensitivity analyses, the weighted seroprevalence adjusted for test performance ranged from 36·3% (95% CI 33·9 – 38·8) to 38·4% (95% CI 35·9 – 41·0)." was already a part of para 3. We believe adding more statements about the sensitivity analyses may not be of any additional benefit in our context. The authors of the STROBE Statement report, "Judgement is needed regarding the level of reporting of such analyses. If many sensitivity analyses were performed, it may be impractical to present detailed findings for them all. It may sometimes be sufficient to report that sensitivity analyses were carried out and that they were consistent with the main results presented. Detailed presentation is more	Changes made in the 'Statistical analysis' sub-section: ***"We used the extremes of the manufacturer-provided 95% CI of the test sensitivity and specificity (upper limit of sensitivity, lower limit of specificity; and lower limit of sensitivity, upper limit of specificity) to report sensitivity analyses."*** The following statement now appears in 'Discussion' section para 4: ***"Using several assumptions about the test sensitivity and specificity to calculate adjusted seroprevalence estimates yielded small differences."***

		appropriate if the issue investigated is of major concern, or if effect estimates vary considerably." ***[Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology. 2007 Nov;18(6):805-35. doi: 10.1097/EDE.0b013e3181577511.]***	

VERSION 4 – REVIEW

REVIEWER	Stoto, Michael Georgetown University
REVIEW RETURNED	01-Sep-2021

GENERAL COMMENTS	As I noted in my earlier reviews, there are fundamentally two problems with this paper. The first is that the results are seriously out of date, and thus not of much interest currently to public health officials or in the long run to scientists. This is not a problem that can be fixed with changes in wording. The potential lasting contribution of this paper is to help sort out the complex relationship between infections, testing, reported cases, and reported deaths, which is a recurring issue in all seroprevalence surveys. The authors seem to have data to help sort this out, but have not taken my suggestions for how to do so. I appreciate the response to other issues raised, but the lack of substantive response to the above two concerns seriously limits the scientific contributions of this paper.

VERSION 4 – AUTHOR RESPONSE

Reviewer comments:
As I noted in my earlier reviews, there are fundamentally two problems with this paper. The first is that the results are seriously out of date, and thus not of much interest currently to public health officials or in the long run to scientists. This is not a problem that can be fixed with changes in wording.

The potential lasting contribution of this paper is to help sort out the complex relationship between infections, testing, reported cases, and reported deaths, which is a recurring issue in all seroprevalence surveys. The authors seem to have data to help sort this out, but have not taken my suggestions for how to do so.
I appreciate the response to other issues raised, but the lack of substantive response to the above two concerns seriously limits the scientific contributions of this paper.
Response:
We appreciate your concerns and the previous and current comments on our manuscript. Your comments have made it possible for us to improve the manuscript quality substantially. In the manuscript, we explicitly mention that the manuscript data may not reflect the current situation in Kashmir. COVID-19 pandemic is very dynamic, and we learn lessons from past experiences. We do not believe that the study has no value since it presents data from Oct-Nov 2020.
Regarding the relationship between infections, reported cases, and deaths, we like to submit the following.
We sincerely appreciate your suggestions during the previous reviews of our manuscript. Accordingly, we carefully studied the two papers (Holtgarve DR et al. and Angulo FJ et al.). and made an effort to act upon the valuable advice.
Holtgrave DR et al. estimate the relative contributions to fatality disparities in terms of differences in SARS-CoV-2 infections, diagnoses, hospitalisations, and death. They have constructed a continuum analogous to the 'care continuum' in the HIV literature. The continuum traces five steps for COVID-19: population size, infection experience with SARS-CoV-2, diagnosis, hospitalisations, and fatality. They discuss the potential reasons for racial and ethnic differences across the continuum and provide essential insights into the factors underlying health disparities in COVID-19 deaths.
Angulo FJ et al. estimate the total number of infections, symptomatic infections, hospitalisations, and deaths in the US as of November 15, 2020. They use data from several seroprevalence surveys conducted across several states in the US to derive "infection underreporting multipliers". Based on CDC Pandemic Planning Scenarios (CDC-PPS), they assume that 60% of infections are symptomatic, 3.4% of symptomatic infections are hospitalised, and 0.65% of infected individuals die. They use this information to estimate the number of infections, hospitalisations, and deaths.
Angulo FJ et al. use the term "infection underreporting multiplier" to quantify the size of underreporting. We use the "number of infections per reported case" instead. They 'estimate' the number of symptomatic infections, hospitalisations, and deaths based on CDC-PPS updated July 1, 2020. In our setting, we do not have appropriate data resources to 'estimate' the number of hospitalisations and deaths, and we believe it is not rational to apply the CDC-PPS figures (even the latest ones, updated March 19, 2021) to our setting without any further research. We, thus, cannot fully reproduce the methods of Angulo FJ et al. in our manuscript. We have, however, reported the total number of reported deaths, which we expect to be much lower than the actual number of COVID-19 deaths in the population because of underreporting.
We made an effort to reproduce the 'continuum' of Holtgrave DR et al., but, as you have pointed out, it does not answer your concern. Hence, we are dropping Figure 5 from our manuscript since it does not add anything significant to the information already presented in the manuscript text. We acknowledge our inability to estimate the true number of hospitalisations and deaths and graphically present the resulting data.