PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study
AUTHORS	Ahmad, Ahsan; Shahzad, Khurram; Masood, Mariumn; Umar, Maida; Abbasi, Fahad; Hafeez, Assad

VERSION 1 – REVIEW

REVIEWER	Puneet Misra All India Inst Med Sci, community medicine
REVIEW RETURNED	20-Jul-2021

GENERAL COMMENTS	good to see this type of work under unity protocol. This is kind of information needed for public health approach

REVIEWER	Mohammed Alghounaim Amiri Hospital, Pediatric Infectious Diseases Specialist and Medical Microbiologist
REVIEW RETURNED	26-Jul-2021

GENERAL COMMENTS	The authors conducted a large randomized cross-sectional seroprevelance study that included individuals from different provinces and regions of Pakistan. The study assesses seropositivity after the first wave of SARS-CoV-2 in Pakistan through a randomized household level sampling technique. The study also assesses different practices (mask-wearing, frequent hand hygiene) with the risk of exposure to SARS-CoV-2. This study may help local authorities as well as countries with a similar structure to infer the need for specific restriction measures and likely immunization planning. However, the study has some limitations, most notably, the selection of high and low- prevalence districts is not clear and may be flawed. Nevertheless, this will not affect overall seroprevalence. Also, some of the association may be affected by recall bias as correctly mentioned by the authors General comments: - The authors keep mentioning “risk-factors” in the manuscript without clearly explaining what risk factors are they evaluating. The way it is mentioned is ambiguous and may refer to risk factors for developing SARS-CoV-2 infection, risk factors for symptomatic disease, and other. Although it becomes clear as I read the manuscript that the authors refer to risk factors for SARS-CoV-2 infection, this need to be clearly mentioned in the abstract and the initial mention of risk factors in the objective and in methods again - Most of the footnotes includes definitions (education, handwashing, … etc) are unnecessary and should be moved to the methods section.

	Abstract (page 2): Line 9: In objectives, it is not clear what the authors refer to in this sentence: “associated risk factors to monitor over time”. Do the authors refer to risk factors to develop disease? Or risk factors for severe disease or factors associated with seropositivity?. The objectives need to be better defined. Line 13: I think it is important to mention how the participants were selected (i.e. random) Line 21: Secondary indicators: please specify at risk of what Line 28: in results, “COVID-19 antibodies” should read SARS-CoV-2 antibodies. COVID-19 is the disease and antibodies are against the virus. Line 35: in conclusion, again it is not clear what the authors refer to by “risk factors for COVID-19”. Do they mean risk factors for developing SARS-CoV-2 infection? Also, the first sentence in the conclusion is a run-on and the meaning is lost. I suggest to rephrase it. Strength and limitation (Pages 2-3): Points 3 to 5 are limitations. They should not be under strengths. Introduction Page 4, line 17: coronavirus disease 2019 is the correct name (no need for “of”) Page 4, line 20: sentence starting with “among various response measures…” contains 2 verbs. Please review it for grammatically accuracy. Page 4 line 37: ELISA is first mentioned here and need to be written in full. Also, this is a run-on sentence. I suggest to start a new sentence at “However, in stark contrast ….” Page 4, line 50: “associated risk factors in the population”. The authors need to be clear on what risk factors they are looking at. I believe they mean risk factors to develop an infection. Methods: Page 5, lines 25-32: the authors defined low and high COVID-19 prevalence by a cut-off of 500 new cases (I assumed they meant new cases/day, this need clarification). However, they didn’t mention if the population of each district was taken into account. Also, with a seroprevalence study, a cumulative prevalence (total number of cases over the period between Feb 26 to June 30, as a percentage of the district’s population) would be more accurate reflection of disease prevalence. This is because serological antibody testing will reflect exposure to SARS-CoV-2 anytime during the pandemic. Not using a cumulative prevalence could be the reason why none of the individual surveyed in Ghotki were positive. Similarly, highest seropositivity was reported in “low” PCR prevalence (table 1). Results: Page 5, line 25: COVID-19 antibodies should be SARS-CoV-2 antibodies. Page 5, lines 27-31: several errors are noted in the text Page 5, line 31: COVID-19 positive should be modified to SARS-CoV-2-infected. Page 7, table 2: the numbers in district prevalence reported does not match table 1. Table 1 shows that 270 and 81 were in low and high prevalence districts, respectively. Table 2 shows the opposite. Also, the authors explain what is the timeline of the symptoms: are those symptoms present at the questionnaire time or previous

	symptoms? Further details regarding the content of the questionnaire should be mentioned in the methods section Page 8, line 3-7: Was the association between the reported findings evaluated? (Handwashing and location, handwashing and gender, and education; similarly with mask use) Page 8, table 4: the reference variable should be mentioned. Also, the variables included in the model need to be clearly stated. Page 8, table 5: should be included in the text of the methods Discussion Page 9, line 45: reference 13 was published in a peer reviewed journal and, thus, should be cited accordingly (https://www.nature.com/articles/s41587-020-0659-0). Also, the authors should highlight that the assay used was evaluated in a small number of SARS-CoV-2 confirmed patients (n=95) Page 9, line 46: the authors provide contradicting arguments. At first, they state that their assay is generally sensitive (up to 100% after 20 days of infection), then, they claim that the prevalence is probably "much higher" due to reduced assay sensitivity within 20 days of infection. Based on the reported cases in Pakistan, the first wave lasted between May to end of July 2020, which is 3 months before the conduction of the study. So, the effect of reduce sensitivity due to early disease is negligible. The authors argument can be correct if they provided that the RDT sensitivity is much lower than an ELISA- or CLIA-based assay. Also, the rate of antibody decay may also contribute in both lower sensitivity and age-specific pattern (https://www.thelancet.com/action/showPdf?pii=S2589-5370%2821%2900182-6) Page 9, line 49: cross reactivity between SARS-CoV-2 Ab testing and other organisms should be viewed with the protein-specific IgG/IgM. Most commercial assays are measuring anti-N or anti-S antibodies, which both have low-level cross-reactivity. The Bioperfectus assay detects both. Page 10, lines 12-16: reduced expression of ACE2 receptors in children is one of many proposed hypotheses. In addition, the role of innate immunity plays a more important role (https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2777743). Page 10, lines 28-39: did the authors observe difference in the IgM positivity pattern between the two tiers that would support their hypothesis? Conclusion: Page 11, line 42: the link between the observed seropositivity and number active cases during the study period is misleading. Seropositivity reflects cumulative cases since the onset of pandemic. This may also will likely reflect on the changes between the two tier-district classifications.

REVIEWER	taoufik Zoubeidi United Arab Emirates University, Statistics
REVIEW RETURNED	10-Oct-2021

GENERAL COMMENTS	The manuscript presents the results of a study to estimate the prevalence of antibodies to Severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) and associated risk factors Pakistan. The adopted sampling design selected one subject of a specific age group from a household. Equal numbers of households were

	selected from a total of 10 districts. The selection of districts, and the selection of subjects of various age groups from the households does not take into account the population size of the districts and the age distribution of subjects within the population. Based on this design, the estimation of the prevalence of COVID-19 antibodies in the whole population may be biased. The authors have pointed to this limitation in the manuscript, but it is important to stress it. The explanation of the observation that the seroprevalence was lower in Tier 1 (districts which had reported high numbers of COVID-19 cases) than in Tier 2 (districts which reported lower COVID-19 cases) needs a more thorough discussion. More information about the population sizes, socio-economic level, availability of healthcare, and any other potentially confounding factors in the participating districts from Tier 1 and Tier 2 would provide a better understanding of this observation. It is not clear how the odds ratios in Table 2 were computed. Were they obtained from a multiple logistic regression, or were they computed directly from the cross-tables? Check the odds ratios under the variable Age in Table 2. Describe the method used to compute the odds ratios in the presented tables.

VERSION 1 – AUTHOR RESPONSE

Serial #	**Comment**	**Explanation/Response**
1.	- Please revise the 'Strengths and limitations' section of your manuscript (after the abstract). This section should contain up to five short bullet points, no longer than one sentence each, that relate specifically to the methods. The results of the study should not be summarised here.	Revised

2.	However, the study has some limitations, most notably, the selection of high and low- prevalence districts is not clear and may be flawed. Nevertheless, this will not affect overall seroprevalence. Also, some of the association may be affected by recall bias as correctly mentioned by the authors	Agreed: The high and low prevalence definition ideally should have been based on percent of cases reported by population in a district. During the initial days of the epidemic, the number of cases reported were quite low. Only 24 districts (out of 136) had reported more than 500 cases. Less than 100 cases were reported by 41 districts. Thus, a strategic decision was made to consider districts reporting more than 500 cases as high prevalence. Which obviously has its limitation. Added to Discussion
3.	The authors keep mentioning “risk-factors” in the manuscript without clearly explaining what risk factors are they evaluating. The way it is mentioned is ambiguous and may refer to risk factors for developing SARS-CoV-2 infection, risk factors for symptomatic disease, and other. Although it becomes clear as I read the manuscript that the authors refer to risk factors for SARS-CoV-2 infection, this need to be clearly mentioned in the abstract and the initial mention of risk factors in the objective and in methods again	Revised
4.	Most of the footnotes includes definitions (education, handwashing, … etc) are unnecessary and should be moved to the methods section.	Footnotes for education and handwashing moved to relevant sections
5.	Line 9: In objectives, it is not clear what the authors refer to in this sentence: “associated risk factors to monitor over time”. Do the authors refer to risk factors to **develop disease?** Or risk factors for severe disease or factors associated with seropositivity?. The objectives need to be better defined.	Revised
6.	Line 13: I think it is important to mention how the participants were selected (i.e. random)	Edited in manuscript
7.	Line 21: Secondary indicators: please specify at risk of what	Revised

8.	Line 28: in results, “COVID-19 antibodies” should read SARS-CoV-2 antibodies. COVID-19 is the disease and antibodies are against the virus.	Revised
9.	Line 35: in conclusion, again it is not clear what the authors refer to by “risk factors for COVID-19”. Do they mean risk factors for developing SARS-CoV-2 infection? Also, the first sentence in the conclusion is a run-on and the meaning is lost. I suggest to rephrase it. (remove?)	Revised
10	Strength and limitation (Pages 2-3): Points 3 to 5 are limitations. They should not be under strengths.	Revised
11	Introduction Page 4, line 17: coronavirus disease 2019 is the correct name (no need for “of”)	Revised
12	Page 4, line 20: sentence starting with “among various response measures…” contains 2 verbs. Please review it for grammatically accuracy.	Revised
13	Page 4 line 37: ELISA is first mentioned here and need to be written in full. Also, this is a run-on sentence. I suggest to start a new sentence at **“However, in stark contrast ….”**	Revised
14	Page 4, line 50: “associated risk factors in the population”. The authors need to be clear on what risk factors they are looking at. I believe they mean risk factors to develop an infection.	Revised
15	Methods: Page 5, lines 25-32: the authors defined low and high COVID-19 prevalence by a cut-off of 500 new cases (I assumed they meant new cases/day, this need clarification). However, they didn’t mention if the population of each district was taken into account. Also, with a seroprevalence study, a cumulative prevalence (total number of cases over the period between Feb 26 to June 30, as a percentage of the district’s population) would be more accurate reflection of disease prevalence. This is because serological antibody testing will reflect exposure to SARS-CoV-2 anytime during the pandemic. Not using a cumulative prevalence could be the reason why none of the individual surveyed in Ghotki were positive. Similarly, highest seropositivity was reported in “low” PCR prevalence (table 1).	These were cumulative cases reported by 30 June. We have corrected this in the manuscript Agreed: The high and low prevalence definition ideally should have been based on percent of cases reported by population in a district. During the initial days of the epidemic, the number of cases reported were quite low. Only 24 districts (out of 136) had reported more than 500 cases. Less than 100 cases were reported by 41 districts. Thus,

		a strategic decision was made to consider districts reporting more than 500 cases as high prevalence. Which obviously has its limitation. (added in discussion) **The estimated seroprevalence was 62 times that of the cases reported by 30th October 2020 in the sampled districts. This points towards a general lack of testing in sampled districts. In Pakistan, testing mostly had been done in symptomatic cases and their contacts. Thus, large pool of sub-clinical infections remained undetected. The variation and low diagnostic testing are likely to be attributable to the gap between seroprevalence and reported cases** **(added in discussion)**
16	Results: Page 5, line 25: COVID-19 antibodies should be SARS-CoV-2 antibodies.	Revised
17	Page 5, lines 27-31: several errors are noted in the text	Revised
18	Page 5, line 31: COVID-19 positive should be modified to **SARS-CoV-2-infected**.	Revised
19	Page 7, table 2: the numbers in district prevalence reported does not match table 1. Table 1 shows that 270 and 81 were in low and high prevalence districts, respectively. Table 2 shows the opposite. Also, the authors explain what is the timeline of the symptoms: are those symptoms present at the questionnaire time or previous symptoms? Further details regarding the content of the questionnaire should be mentioned in the methods section	a) Results are correct in both tables and labels are corrected in table 2. b) Timeline of symptoms was past 3 months. c) The questionnaire captured the information of socio-demographic variables, medical and symptoms history, preventive

		behaviours, complications and history of recent death in family followed by Rapid Diagnostic Test (RDT) results performed by a trained phlebotomist.
20	Page 8, table 4: the reference variable should be mentioned. Also, the variables included in the model need to be clearly stated.	a) Reference category is mentioned in the footnote of Table 4. b) Model variables are explained in the methods section.
21	Page 8, table 5: should be included in the text of the methods	Removed table 5 and added in the methods section.
22	Discussion Page 9, line 45: reference 13 was published in a peer reviewed journal and, thus, should be cited accordingly (https://www.nature.com/articles/s41587-020-0659-0). Also, the authors should highlight that the assay used was evaluated in a small number of SARS-CoV-2 confirmed patients (n=95)	Reference Revised Unable to find the mentioned n=95 in the mentioned reference (https://www.nature.com/articles/s41587-020-0659-0).) or elsewhere in the manuscript.
1.	Page 9, line 46: the authors provide contradicting arguments. At first, they state that their assay is generally sensitive (up to 100% after 20 days of infection), then, they claim that the prevalence is probably “much higher” due to reduced assay sensitivity within 20 days of infection. Based on the reported cases in Pakistan, the first wave lasted between May to end of July 2020, which is 3 months before the conduction of the study. So, the effect of reduce sensitivity due to early disease is negligible. The authors argument can be correct if they provided that the RDT sensitivity is much lower than an ELISA- or CLIA-based assay. Also, the rate of antibody decay may also contribute in both lower sensitivity and age-specific pattern (https://www.thelancet.com/action/showPdf?pii=S2589-5370%2821%2900182-6)	The comment is agreeable to possibly connotate, that the sensitivity is 100% while the authors mention limitations of sensitivity to be actually/possibly less. It is submitted, that given that first wave was during May-Jul 2020..the 2nd wave was being experienced when this survey was done in the country. So it is plausible to state, that the sensitivity for those who experienced COVID-19 in the first wave (which cannot be actually determined) may have had more accurate results on RDT; while those who might have experienced the illness in a time frame closer to the

		conduct of survey might have less accurate results. Since the study, did not map the actual time of occurrence of disease, hence the possibility is explained in the discussion session. (Hence No further changed made in page 9, line 46)
2.	Page 9, line 49: cross reactivity between SARS-CoV-2 Ab testing and other organisms should be viewed with the protein-specific IgG/IgM. Most commercial assays are measuring anti-N or anti-S antibodies, which both have low-level cross-reactivity. The Bioperfectus assay detects both. (REMOVE LINE)	Removed after discussion among authors
3.	Page 10, lines 12-16: reduced expression of ACE2 receptors in children is one of many proposed hypotheses. In addition, the role of innate immunity plays a more important role (https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2777743). (ADD REFEERNCE)	Revised, Reference added
4.	Page 10, lines 28-39: did the authors observe difference in the IgM positivity pattern between the two tiers that would support their hypothesis? (re-run?)	Yes. IgM positivity level were lower among tier 1 districts as compared to tier 2 districts and their association was also significant.
5.	Conclusion: Page 11, line 42: the link between the observed seropositivity and number active cases during the study period is misleading. Seropositivity reflects cumulative cases since the onset of pandemic. This may also will likely reflect on the changes between the two tier-district classifications.	Agreeable. We have modified accordingly.
6.	The adopted sampling design selected one subject of a specific age group from a household. Equal numbers of households were selected from a total of 10 districts. The selection of districts, and the selection of subjects of various age groups from the households does not take into account the population size of the districts and the age distribution of subjects within the population. Based on this design, the estimation of the prevalence of COVID-19 antibodies in the whole population may be biased. The authors have pointed to this limitation in the manuscript, but it is important to stress it.	Added as Paragraph in the Discussion Section

7.	The explanation of the observation that the seroprevalence was lower in Tier 1 (districts which had reported high numbers of COVID-19 cases) than in Tier 2 (districts which reported lower COVID-19 cases) needs a more thorough discussion. More information about the population sizes, socio-economic level, availability of healthcare, and any other potentially confounding factors in the participating districts from Tier 1 and Tier 2 would provide a better understanding of this observation.	High reported seroprevalence in Tier-1 vs Tier-2 districts could also be due to confounding factors. The average population size of the districts in Tier-1 districts was 2.9million compared to 0.9 million in Tier-2. Similarly, the average population density in Tier-1 districts was more than double of that in Tier-2 districts (852 vs 348 persons per square kilometre) Added as an additional paragraph in discussion
8.	It is not clear how the odds ratios in Table 2 were computed. Were they obtained from a multiple logistic regression, or were they computed directly from the cross-tables? Check the odds ratios under the variable Age in Table 2. Describe the method used to compute the odds ratios in the presented tables.	Univariate logistic regression of seropositivity with age, location, gender, district tiers, symptoms and covid contact was computed. Description has been added in the methods section.

VERSION 2 – REVIEW

REVIEWER	Mohammed Alghounaim Amiri Hospital, Pediatric Infectious Diseases Specialist and Medical Microbiologist
REVIEW RETURNED	24-Jan-2022

GENERAL COMMENTS	- Abstract: I suggest that the authors just write the primary and secondary outcome measures that were measured in the abstract. Details can be mentioned in the manuscript. - There is a problem with the citation style. It includes regular and square brackets: i.e. “(1)]” as well as some references were not captured by the citation software (i.e. (Error! Reference source not found.)

REVIEWER	taoufik Zoubeidi United Arab Emirates University, Statistics
REVIEW RETURNED	05-Feb-2022

GENERAL COMMENTS	MY remarks have been addressed by the authors.

VERSION 2 – AUTHOR RESPONSE

Comments- Reviewer: 2	Response
Abstract: I suggest that the authors just write the primary and secondary outcome measures that were measured in the abstract. Details can be mentioned in the manuscript.	Revised Abstract
There is a problem with the citation style. It includes regular and square brackets: i.e. “(1)]” as well as some references were not captured by the citation software (i.e. (Error! Reference source not found.)	- Revised Citations - Removed Errors