BMJ Open

COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055381
Article Type:	Original research
Date Submitted by the Author:	13-Jul-2021
Complete List of Authors:	Ahmad, Ahsan ; Health Services Academy Shahzad, Khurram; Health Services Academy Masood, Mariumn; Health Services Academy Umar, Maida; Health Services Academy Abbasi, Fahad; Health Services Academy Hafeez, Assad; Health Services Academy
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, Epidemiology < TROPICAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study

Corresponding Author: Dr. Mariumn Masood, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan. Email: msmasood148@gmail.com

Author Information

1. Dr. Ahsan M Ahmed, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
ahsen@cgph.org.pk

2. Dr. Khurram Shahzad, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
baig.khurram@gmail.com

3. Dr. Mariumn Masood, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
msmasood148@gmail.com

4. Ms. Maida Umar, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
maidakhawja@gmail.com

5. Dr. Fahad Abbasi, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
idrfahadabbasi@gmail.com

6. Dr. Assad Hafeez, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
az10@hotmail.com

Word Count: 3008 words

Ethics Approval: from the Institutional Ethical Review Committee of Health Services Academy, Islamabad, Pakistan (Location: Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan)

ABSTRACT

Objectives This study adapted The World Health Organization's 'Unity Study' protocol to estimate the population prevalence of antibodies to Severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) and associated risk factors to monitor its patterns over time.
Design This population-based, age-stratified cross-sectional study was conducted in households (HH).
Participants 4,998 households out of 6,599 consented (one individual per household). 51% were male participants. All ages and sexes were eligible. (Exclusion Criteria: contraindications to venipuncture. However no such case was encountered).
Primary and Secondary Outcome Measures
Following were the Planned Outcome Measures
Primary Indicators: i) Seroprevalence (population and age-specific); ii) Asymptomatic fraction (proportion of cases that are asymptomatic)
Secondary Indicators**:** i) Population groups most at risk.
Following were the Measured Outcome Measures. These were different from the planned indicators (i.e. two out of the three planned indicators were measured) due to operational reasons and time constraints.
Primary Indicators: i) Seroprevalence (population and age-specific)
Secondary Indicators: Population groups most at risk.
Results Overall seroprevalence of COVID-19 antibodies was 7.1%. 6.3% of individuals were IgG positive while IgM positivity was 1.9%. The seroprevalence among different age groups ranged from 3.9% (0-9 years) to 10.1% (40-59 years). Seroprevalence in districts ranged from 0% (Ghotki) to 17% (Gilgit). A history of contact with a confirmed COVID-19 case, being an urban resident and mask use were key risk factors for sero-positivity.
Conclusions This survey provides useful estimates for seroprevalence in the general population and information on risk factors for COVID-19 at the beginning of the second wave, with the seropositivity results corresponding with the phase of the pandemic in the country. It is premised that similar studies need to be replicated at the population level on a regular basis to monitor the disease and immunity patterns related to COVID-19.

Strengths and Limitations of the Study
Strengths:

- The study provides seroprevalence estimates from a large sample size of about 5000 individuals, with random selection at each stage i.e. district, Union Council, Villages (for rural Union Councils) households, and individuals, and representation from all four provinces and both regions of the country
- There is an almost equal representation from both the sexes; and the rural and urban samples were in accordance with the proportion of urban: rural Union Councils of each district.
- Although RDTs (that were used to estimate the seroprevalence) were selected for their advantage over other methods due to their ease-of-use in the field, the seroprevalence estimates may have been affected by the low sensitivity of RDTs, causing underestimation of the seroprevalence values.

- The estimates for association of seropositivity with some of the studied risk factors may have been affected by biases including recall bias and social desirability bias, especially among the self-reported risk factors.
- The seroprevalence estimates provided by this study may be interpreted with caution as an estimated value for the general population, particularly since the age structure of the study population differs from the demographic distribution of the population in the country.

INTRODUCTION

The Coronavirus disease of 2019 (COVID-19), originating from Wuhan, China was declared a pandemic by the World Health Organization (WHO) on the 11th of March, 2020.[1] Pakistan reported its first case on February 26th, 2020, witnessing its first peak during July, 2020.[2] Among various response measures undertaken in Pakistan have been the conduction of research studies to inform response measures to COVID-19 and to enable a better understanding of its epidemiology and spread. This includes partaking in the WHO's global research initiatives such as the 'WHO Unity Studies' through the conduction of national seroprevalence studies.[3]These study protocols have also been adapted around the world at various geographic levels to provide contextual data on the evolving pandemic.

Numerous seroprevalence studies have been conducted around the globe since the onset of the pandemic, with more evidence from large-scale nationwide studies being reported as the situation evolves, and the results showing wide contextual variations.[4]In neighboring countries, Iran reported a high seroprevalence of 17.1% in a large study conducted across 17 provinces during April to June 2020, although the results were based on, and were from a much earlier phase of the pandemic.[5] India also conducted a national seroprevalence survey (for IgG) in adults during the same period utilizing ELISA, however, in stark contrast to Iran it reported markedly lower seroprevalence of 0.7% in its study population.[6] Studies in the other regions also depict varying patterns of seroprevalence depending on the timelines of the pandemic in their respective countries.[7-9]One of the largest seroprevalence studies around the globe in England has reported a decline in seroprevalence estimates of IgG by 26.5% between June and September 2020, from 6.0% to 4.4% by lateral flow immunoassay (LFIA) tests.[10]

This study in Pakistan is part of the global 'WHO Unity Studies' initiative with the main objective of estimating the seroprevalence of IgG and IgM antibodies to the 'SARS-CoV-2' coronavirus in the general population in the country. This study will not only provide data regarding the exposure of the general population to COVID-19, but would also shed light on some of the associated risk factors in the population. Overall, we expect that the estimates would provide us with ample evidence to gauge the population-level scenario of COVID-19 in the country as well as provide insights into other epidemiological aspects of the disease and its risk factors.

METHODS

Patient and Public involvement Statement: Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

This population-based, age-stratified cross-sectional study, was conducted at the level of households (HH) between October 21st and November 8th, 2020, with Ethics Approval from the Institutional Ethical Review Committee of Health Services Academy, Islamabad.

This was a nationwide study in the four provinces (Punjab, Sindh, Baluchistan and Khyber Pakhtunkhwa/KP) and two regions (Azad Jammu & Kashmir (AJK) and Gilgit Baltistan (GB)) of Pakistan. Individuals of all ages and genders were eligible to participate.

Sample size calculation was done using Open Source Epidemiologic Statistics for Public Health [a] using multistage sampling design (with expected prevalence of 11% which were preliminary estimates from a previous national level study;[11]) difference between upper and lower limit of the interval estimate as 1.25% (0.75% on either side); and a design effect of '2'). The resultant sample size of 4,803 was rounded off to 5,000 and distributed equally among the ten districts under study as well as among all age brackets.

The study was conducted in ten districts of the country according to the following criteria: one high and one low prevalence district each was selected from provinces; and the highest-prevalence district each was selected from regions AJK and GB. In Pakistan's context, high prevalence was taken as more than 500 COVID-19 cases (Tier 1); and low prevalence as less than 500 cases (Tier 2) by the cut-off date of June 30th, 2020. The sample from provinces had thus included equal representation from high and low prevalence districts.

After district selection, Union Council (UC) selection was done with the aim of recruiting 500 households/participants per districts (25 participants defined as one cluster). UCs were randomly selected from each district; the selection was in accordance with the "Urban UC: Rural UC" ratio for each district. Where there were less than 20 UCs in a district, the number of clusters were increased to make the total equal to 500 participants per district. Systematic random sampling was employed for the next stage i.e. household selection. Thereafter, one individual was randomly selected from each consenting household in line with the age distribution of the study. Each cluster of 25 had five participants from each age group. In this way 25 HH/individuals were recruited for each cluster, with random selection at different stages (as described) aiming to reduce potential sources of selection bias.

Data collection for the survey included on-the-spot recruitment of households in the selected localities. After obtaining informed consent, each participant(one individual randomly selected from each household) was asked to provide information to the enumerator to fill a pre-tested questionnaire (adapted from the WHO Unity Studies protocol [3]); this was followed by a Rapid Diagnostic Test/RDT performed by a trained phlebotomist. RDT was performed with 'Bioperfectus' kits for IgG/IgM, and the results were provided to each participant on the

[a] https://www.openepi.com/Menu/OE_Menu.htm

spot(within a few minutes). Data analysis for the survey (including descriptive statistics and multivariate regression modeling) was performed using SPSS version 23. Prior to the survey, data collectors and field teams underwent training, which included PPE usage for infection prevention and control. Monitoring visits were also conducted during the data collection period to enhance data quality and results.

RESULTS

During field work, recruitment was continued till the target sample was achieved. A total of 6,599 households were reached for on-the-spot participation, of which 1,601 (24%) did not provide consent to participate. A total of 4,998 households (with one individual per household) consented to participate in the study across the ten selected districts of Pakistan. Individuals of all ages and genders were eligible. The proportion of males in the recruited participants was 51%. The mean age of males (31.7 years) was similar to that of females (32.8 years). Almost two thirds (62%) of individuals were recruited from rural areas. Male to female ratio of sampled individuals was similar for rural and urban areas.

Information on seropervalence was available for all 4998 participants with no missing values. The overall seroprevalence of COVID-19 antibodies was 7.1%. Almost 6.3% of individuals were IgG positive while IgM positivity was 1.9%. Seroprevalence in districts ranged from 0.0% (Ghotki) to 17.0% (Gilgit). Most of the districts reported a range of 6% to 9% (**Error! Reference source not found.**).The seroprevalence among different age groups ranged from 4% (0-9 years) to 10% (40-59 years). (**Error! Reference source not found.**). A total of 4% reported to have had contact with a COVID-19 positive individual.

Table 1 : Seroprevalence at District Level

District	Reported prevalence based on RT-PCR[a]	Positive (%)	Negative (%)
Quetta	High	16(3.2)	484(96.8)
Mardan		31(6.2)	470(93.8)
Rawalpindi		34(6.8)	467(93.2)
Ghotki		0 (0.0)	498(100)
Muzaffarabad	Low	39(8.6)	460(92.2)
Gilgit		85(17.0)	415(83.0)
Sibbi		24(4.8)	474(95.2)
Abbottabad		42(8.4)	460(91.6)
Lodhran		43(8.6)	457(91.4)
Jacobabad		37(7.4)	462(92.6)

[a] District reporting more than 500 cases as of 30 June 2020 were categorized as high prevalence (tier-1) and those reporting less than 500 as low prevalence

Table 2 Seroprevalence by age, place, district tier, gender and (significant)symptoms

Variables		Positive (%)	Negative (%)	Odds ratio (95% CI)	p-value
Seroprevalence for COVID-19		351(7.1)	4647(93.0)		
Gender (*n*=4,998)	Male	178(6.9)	2394(93.1)	0.97(0.78-1.20)	0.879
	Female	173(7.1)	2253(92.9)		
Age (years) (*n*=4,997)	0 – 9	38(3.9)	935(96.1)		
	10 – 19	46(4.5)	973(95.5)	0.43(0.29-0.63)	0.000
	20 – 39	80(7.8)	950(92.2)	0.49(0.34-0.71)	0.000
	40 – 59	101(10.1)	899(89.9)	0.88(0.64-1.21)	0.390
	60+	85(8.7)	890(91.3)	1.18(0.87-1.59)	0.309
Location (*n*=4,998)	Urban	154(8.1)	1741(91.9)	1.30(1.04-1.62)	0.014
	Rural	197(6.3)	2906(93.7)		
District prevalence (reported cases) (*n*=4,998)	High prevalence (tier-2)	270(9.0)	1919(96.0)	2.34(1.81-3.02)	0.000
	Low prevalence (tier-1)	81(4.1)	2728(91.0)		
Contact with COVID-19 positive case (*n*=4,971)	Yes	25(13.9)	155(86.1)	2.34(1.81-3.03)	0.001
	No	292(6.7)	4041(93.3)		
Sore throat (*n*=4,987)	Yes	118(10.6)	998(89.4)	1.86(1.46-2.32)	0.000
	No	233(6.0)	3638(94)		
Fatigue (*n*=4,991)	Yes	43(10.9)	351(89.1)	1.706(1.22-2.39)	0.002
	No	308(6.7)	4289(93.3)		
Joint ache (*n*=4,988)	Yes	65(10.7)	541(89.3)	1.734(1.31-2.30)	0.000
	No	284(6.5)	4.98(93.5)		
High grade fever (*n*=4,986)	Yes	123(8.6)	1313(91.4)	1.365(1.09-1.72)	0.008
	No	228(6.4)	3322(93.6)		
Cough (*n*=4,993)	Yes	98(8.5)	1055(91.5)	1.317(1.03-1.67)	0.026
	No	253(6.6)	3587(93.4)		
Runny nose (*n*=4,987)	Yes	97(8.5)	1044(91.5)	1.32(1.03-1.7)	0.026
	No	253(6.6)	3593(93.4)		

The use of preventative behaviors was also studied. The use of face-masks while going out in public was reported to be 63%. Mask use was similar in urban and rural areas (63%). It was highest in 20-59 years age group (68%), while 60+ group reported relatively less use (49%). Mask use increased incrementally with education, from 36% in those non-educated to 82% in individuals

above matric. Handwashing[b] was reported relatively less compared to mask use (39%). Handwashing was higher in urban (44%) areas; among females (43%); and increased with education, being highest in individuals above matric[c] (58%) (**Error! Reference source not found.**).

Table 3 Mask use and handwashing practices by socio-demographic characteristics

Variables		Mask use (%)			Handwashing (%)[a]		
		Yes	No	Total(%)	Yes	No	Total(%)
Overall		3,128(62.6)	1,844(37.1)		1,946(38.9)	3,052(61.1)	
Location	Urban	1,185(62.9)	699(37.1)	1,884(37.8)	839(44.3)	1,056(55.7)	1,895(37.9)
	Rural	1,943(62.9)	1,145(37.1)	3,088(62.1)	1,107(35.7)	1,996(64.3)	3,103(62.1)
District Tiers	Tier-1	1,121(56.5)	864(43.5)	1,985(39.9)	762(38.1)	1,238(61.9)	2,000(40.0)
	Tier-2	2,007(67.2)	980(32.8)	2,987(60.0)	1,184(39.5)	1,814(60.5)	2,992(60)
Gender	Male	1,654(64.6)	907(35.4)	2,561(51.5)	905(35.2)	1,667(33.4)	2,572(51.5)
	Female	1,474(61.1)	937(38.9)	2,411(48.5)	1,041(42.9)	1,385(27.7)	2,426(48.5)
Age	0-19	576(59.5)	392(40.5)	968(24.5)	264(27.1)	709(72.9)	973(24.5)
	20-59	1,364(67.5)	657(32.5)	2,021(51.5)	934(46.0)	1,096(54.0)	2,030(51.0)
	60+	476(49.1)	50.9(50.9)	970(24.5)	382(39.2)	593(2398)	975(24.5)
Education	No education	412(36.0)	731(39.0)	1,143(36.9)	326(28.4)	822(71.6)	1,148(36.9)
	Primary	194(61.0)	124(24.8)	318(10.3)	137(42.9)	182(57.1)	319(10.3)
	Matric	491(75.2)	162(18.4)	653(21.2)	307(46.5)	353(53.5)	660(21.2)
	Above matric	797(81.6)	180(64.0)	977(36.9)	573(58.4)	408(13.1)	981(31.6)

[a] washing hands at least six times with soap and water for 20 seconds in last 24 hours

Symptoms during previous months were inquired to look for possible association with COVID-19 seropositivity. Total of 23 symptoms were inquired relating to multiple systems. The symptoms shown in **Error! Reference source not found.** were significantly higher in seropositive individuals. Sore throat, fatigue, and joint aches were strongly associated with seropositivity. Among COVID-19 seropositive individuals, 68% had at least one symptom in last two months, while 32% reported to be completely asymptomatic during this period.

Almost 24% of individuals reported having at least one comorbidity. Hypertension was reported the most (18%), followed by diabetes (5%) and chronic kidney disease (2%). The reported

[b] Individuals were inquired about number of times they had washed their hands with soap for 20 seconds. The variable was categorized into those washing hands at least six times and those less than six times.
[c] Ten years of education

occurrence of heart disease (1%) and asthma (2%) was relatively lower. Reported prevalence of at least one comorbidity increased with age with maximum being reported for 60+ age group (54%), followed by 40-59 years age group (40%).

Multivariate logistic regression analysis was performed to identify factors associated with seropositivity (Table 4). Urban residents were more likely to test positive for COVID-19 antibodies than rural residents (OR 1.29, 95% CI 1.04 - 1.61). Individuals aged 20 and above were about twice as likely to be seropositive than those who were 0-9 years old. Odds of seropositivity were also high among individuals who did not wear face mask (OR 1.54, 95% CI 1.20 – 1.975) and in those who reported contact with COVID-19 person (OR 1.81, 95% CI 1.16 – 2.83).

For any discrepancy and to ensure completeness of data, field teams were contacted and cross-checked. All cases are included in the analysis and missing data upto 1% were considered acceptable because of low occurrence. Table 5 describes other analysis details.

Table 4 Factors associated with seropositivity

Variables		**b**	**S.E.**	**Adjusted OR (95% CI)[a]**	**p-value**
Location	Urban	0.256	0.113	1.29(1.035-1.612)	0.024
	Rural				
Age (years)	0 – 9				
	10 – 19	0.112	0.225	1.11(0.7-1.737)	0.618
	20 – 39	0.650	0.204	1.92(1.285-2.854)	0.004
	40 – 59	0.975	0.197	2.65(1.803-3.899)	0.000
	60+	0.894	0.202	2.45(1.646-3.630)	0.000
Mask use	Yes	0.434	0.126	1.54(1.205-1.975)	0.001
	No				
Contact with COVID-19 positive case	Yes	0.596	0.227	1.81(1.163-2.831)	0.009
	No				
Constant		-3.599			

[a] Hosmer-Lemshow goodness-of-fit test: chi-square value = 8.322, P = 0.403

Table 5: Additional details on methods used for Multivariate Regression Analysis

Aspect of Analysis	Description
Model for multivariate binary logistic regression	$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_i X_i \quad (1)$$ where p is the probability that an individual is seropositive, β_0 is the intercept, β_i are the coefficients and X_i are the independent variables including location, age, mask use and contact with COVID-19 positive person.

Variable and Model Selection	A forward-stepwise process was utilized for the selection of significant variables in the final model. In the analysis, some interaction terms were considered and not included in the final model because they were not statistically significant. The selection of multivariate logistic model was based on Homer-Lemeshow goodness of fit, biological interpretability and statistical significance.
Significance level or alpha	0.05

DISCUSSION

The survey through the use of Unity Studies' age-stratified approach, estimated the national seroprevalence of IgG and IgM antibodies for COVID-19 to be 7.1% in Pakistan based on RDT testing. Among the included districts, the highest prevalence was observed for Gilgit, followed by Lodhran and Muzaffarabad (Table 1). This study was initiated in the last week of October 2020 and field activities were completed by second week of November 2020. This was a time frame, when the first wave of the pandemic was considered to have largely subsided, the new number of cases per day was markedly lower, and there was a threat of a second wave of the pandemic in the forthcoming winter months of November onwards. During this time frame < 1000 confirmed cases per day were being recorded, in the backdrop of the highest daily number of cases (6,825) that had been reported on 13th June, 2020.[2]

Although other tests such as ELISA (Enzyme-linked immunosorbent assay) offer greater accuracy than RDTs in terms of antibody detection,[12] due to practical and operational issues, RDTs were opted for this large-scale population-based study to estimate the seroprevalence of antibodies to COVID-19, with the results for on-the-spot testing in the field available within twenty minutes. Some researchers using similar kits to those used in this survey reported the sensitivity (to detect IgG/IgM)to range between 41% (at 1-5 days since symptom onset in patients positive by RT-PCR) to 100% (at > 20 days since symptom onset in patients positive by RT-PCR); while the reported the specificity when compared with PCR was 95% .[13] This time-dependent sensitivity of RDTs to detect SARS-Cov-2 antibodies has also been noted elsewhere.[14] Due to the widely varying sensitivity of the testing method, it is likely that the actual seroprevalence may have been much higher than this study's estimates. In contrast, one of the other limitations of RDT testing is considered to be cross-reactivity to other antibodies (particularly those against other Coronaviruses),[15, 16] which may also have potentially affected the results-albeit by overestimating the prevalence to some extent.

About one third of seropositive individuals had reported to have experienced symptoms during the past two months. Six out of the studied 23 symptoms experienced during the past two months were found to be significantly associated with seropositivity in the univariate analysis, most of which

were respiratory/pharyngeal symptoms including sore throat, shortness of breath, cough and runny nose. High grade fever, joint aches and fatigue were the three generalized symptoms associated with seropositivity. Although the symptoms were self-reported with a possibility for recall bias, information on symptoms was obtained before the testing was done hence it is likely that any misclassification may have been non-differential.

Similar to what has been reported in some other national studies (including a large-scale national-level household study by Pollán et al in Spain), gender was not found to be significantly associated with seropositivity for COVID-19 antibodies.[8, 9, 17] The association of age with seropositivity increased with age until 59 years, and declined slightly in those above this age bracket. As observed previously, nasal gene expression of angiotensin-converting enzyme 2 (ACE2) has been postulated to be responsible for this age-related pattern, and children have been reported to be less susceptible to contracting COVID-19 than adults potentially due to the protective effect of lower levels of this enzyme in their nasal epithelium, as also reported elsewhere.[18, 19]

The risk of seropositivity doubled in those with a history of exposure to a diagnosed COVID-19 patient and was found to be statistically significant. A study in Italy reported an even higher Odds ratio of 2.5 in those who had previous contact with a case. [20] Association of COVID-19 with household contact with known cases of COVID-19 has also found to be significant in other prevalence studies.[21].

Differences in seroprevalence between Tier 1 (districts considered to be high transmission areas by the end of June 2020) and Tier-2 districts (districts considered to be low transmission areas at the end of June 2020), were significant at about 4.9%. However, it is important to note that the seroprevalence was lower in Tier 1 districts in this study. This pattern suggests that the transmission scenarios would have evolved in three-four months since the chosen cut-off date, and the areas earlier considered to be higher transmission at the end of June (i.e. Tier 1 districts) may have now become areas of relatively lower transmission, and vice versa. The said difference is understood to be possibly due to a higher proportion of population in previously high risk districts, having possibly experienced the exposures, with possibility of reduction in IgM levels across a time span of 90 or more days (i.e. the time span between June and October 2020).

Urban and rural areas were sampled from each district in accordance with the urban to rural ratio for that particular district. Overall, urban areas reported a higher prevalence (8.1%) than rural areas (6.3%), and urban residents were more likely to be seropositive for COVID-19 than rural residents (Table 4). While some researchers have argued that larger city sizes tend to have higher attack rates, [22] other studies have gone further so as to report that while urban areas do have a propensity for earlier outbreaks than rural areas, population density is not significantly associated with COVID-19 cases.[23] Our results seem to align with the former, i.e. showing a significant association of urban residence with COVID-19 seropositivity.

Among studied behaviors, use of masks was reported by about two-thirds of the study participants with similar values in rural and urban areas. Mask use was found to be significantly linked with seropositivity in logistic regression in our study (Table 4), in line with wide-ranging evidence.(1, 24, 25) On the other hand, hand-washing(at least six times per day) had a lower prevalence (39%)

in the study population and was not found to be a significant risk factor in this study, although other researchers have reported protective benefits of hand hygiene [26]. Since these have both been recommended preventive behaviours during the pandemic, the results may have been affected by Social Desirability Bias[27] in the study population, causing over-reporting of these behaviours. In such a case, any true association of these behaviours with seropositivity for COVID-19 may have been masked, if they do indeed influence seropositivity for SARS-CoV-2.

Various studies to estimate the seroprevalence of COVID-19 have been conducted in the country at different scales. For example, preliminary results from a national sero-prevalence study that was conducted during July 2020 reflected a prevalence of 11.2%.[11] Seroprevalence studies at a smaller scale in the country have also been conducted in Karachi city and Islamabad Capital Territory(ICT), with the former reporting figures of 9.7% and 15.1% in 'low-transmission' and 'high-transmission' areas in the city, respectively,[28] and the latter reporting an overall seroprevalence of 14.5%.[29, 30] Results from the aforementioned sub-national surveys differ from those of the current study in which the seroprevalence was found to 4.1% in Tier-1 districts and 9.2% in Tier-2 districts, respectively. This may have been due the differing time frames of the studies in Karachi and ICT, which were conducted during the first wave of COVID-19 in the country; while the current study was conducted during October-November 2020, when the first wave of the pandemic had largely subsided.

Overall, it is likely that the seroprevalence estimates may have been affected by low sensitivity of the testing methods(causing underestimation of the values); and the estimates for association of seropositivty with the studied risk factors may have been affected by biases including recall and social desirability bias among the study population. The seroprevalence estimates provided by this study may be interpreted with caution as an estimated value for the general population, particularly since the age structure of the study population differs from the demographic distribution of the population in the country.

CONCLUSIONS

Notwithstanding the limitations of the study, this survey provides useful prevalence estimates as well as information on risk factors. Conducted early during the second wave with newly reported cases still relatively low, the seropositivity results correspond with the phase of the pandemic in the country. The results also show that the youngest age groups have the lowest proportion of sero-positivity as compared to those aged 40 years and above. Interestingly, Tier 1 districts (considered to be high risk based on the number of PCR test based confirmed cases by the end of June 2020) reflected lower prevalence as compared to Tier 2 districts which may perhaps be depicting reversing patterns at the population level. A history of contact with a confirmed COVID-19 case, being an urban resident and mask use were key risk factors for seropositivity. Keeping view of these findings, it is premised that similar studies need to be replicated at the population level on a regular basis to monitor the disease and immunity patterns related to COVID-19.

List of Abbreviations

COVID-19- The Coronavirus disease of 2019

WHO- World Health Organization

ELISA - Enzyme-linked immunosorbent assay

RDT - Rapid Diagnostic Test

SARS-CoV-2 - Severe acute respiratory syndrome coronavirus 2

RT-PCR - Reverse transcription polymerase chain reaction

ACE2- Angiotensin-converting enzyme 2

HH- Household

OR- Odds Ratio

IRB - Institutional Review Board

KP - Khyber Pakhtunkhwa

AJK- Azad Jammu & Kashmir

GB- Gilgit Baltistan

UC- Union Council

LFIA - Lateral flow immunoassay

- **Competing interests**

  The authors declare that they have no competing interests

- **Funding**

  This study was funded by Health Services Academy (HSA), Islamabad which is affiliated with the Ministry of National Health Services Regulations and Coordination, Government of Pakistan. HSA was involved in all aspects of the survey including study design/technical supervision; collection, analysis and interpretation of the data; in the writing of the manuscript; and in the decision to submit the paper for publication.

  Funding/grant award number: F-11-2020-WHO/DAi/HSA

- **Authors' contributions**

AH and AA supervised the overall process; AA contributed to the analysis, manuscript writing and monitoring of data collection; KS led the coordination of all survey activities including design, training, administrative aspects and data analysis; led the process of preparing tools for data collection; contributed to training of data collectors and monitoring of the data collection process, and to the analysis and reporting of Results; MM supported preparation of data collection tools, contributed to training of data collectors, drafted the Household sampling methodology, conducted household refusal data analysis, contributed to monitoring of data collection, conducted literature reviews to analyse epidemiological data from the survey, and prepared the initial manuscript draft; MU contributed to data management, analysis and reporting. FA supported preparation of data collection tools, contributed to monitoring of data collection and training of data collectors. All authors reviewed and approved the manuscript.

- **Acknowledgements**

Not Applicable

References

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 2020 [cited 2020 22 December]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
2. Government of Pakistan. COVID-19 Dashboard 2020 [cited 2020 22 December]. Available from: http://covid.gov.pk/stats/pakistan?locale=en.
3. World Health Organization. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigation Protocols 2020 [cited 2020 22 December]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations.
4. Arora RK, Joseph A, Van Wyk J, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. The Lancet Infectious Diseases. 2020.
5. Poustchi H, Darvishian M, Mohammadi Z, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. The Lancet Infectious Diseases. 2020.
6. Murhekar MV, Bhatnagar T, Selvaraju S, et al. Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020. Indian Journal of Medical Research. 2020;152(1):48.
7. Xu X, Sun J, Nie S, et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. Nature Medicine. 2020:1-3.
8. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. The Lancet. 2020;396(10250):535-44.
9. Herzog S, De Bie J, Abrams S, et al. Seroprevalence of IgG antibodies against SARS coronavirus 2 in Belgium: a prospective cross-sectional study of residual samples. medRxiv. 2020.
10. Ward H, Cooke G, Atchison CJ, et al. Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. MedRxiv. 2020.
11. Dawn News. 11pc Pakistanis have developed protective immunity: study 2020 [cited 2020 22 December]. Available from: https://www.dawn.com/news/1575607/11pc-pakistanis-have-developed-protective-immunity-study.
12. Kontou PI, Braliou GG, Dimou NL, et al. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. Diagnostics. 2020;10(5):319.
13. Whitman JD, Hiatt J, Mowery CT, et al. Test performance evaluation of SARS-CoV-2 serological assays. MedRxiv.
14. Rashid ZZ, Othman SN, Samat MNA, et al. Diagnostic performance of COVID-19 serology assays. The Malaysian Journal of Pathology. 2020;42(1):13-21.
15. Ma Z, Li P, Ji Y, et al. Cross-reactivity towards SARS-CoV-2: the potential role of low-pathogenic human coronaviruses. The Lancet Microbe. 2020;1(4):e151-e.
16. Nuovo G, Tili E, Suster D, et al. Strong homology between SARS-CoV-2 envelope protein and a Mycobacterium sp. antigen allows rapid diagnosis of Mycobacterial infections and may provide specific anti-SARS-CoV-2 immunity via the BCG vaccine. Annals of diagnostic pathology. 2020;48:151600.
17. Lai C-C, Wang J-H, Hsueh P-R. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: An up-to-date review. International Journal of Infectious Diseases. 2020.
18. Patel AB, Verma A. Nasal ACE2 levels and COVID-19 in children. Jama. 2020;323(23):2386-7.

19. Bunyavanich S, Do A, Vicencio A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. Jama. 2020.
20. Vena A, Berruti M, Adessi A, et al. Prevalence of antibodies to SARS-CoV-2 in Italian adults and associated risk factors. Journal of clinical medicine. 2020;9(9):2780.
21. Menachemi N, Yiannoutsos CT, Dixon BE, et al. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. Morbidity and Mortality Weekly Report. 2020;69(29):960.
22. Stier A, Berman M, Bettencourt L. COVID-19 attack rate increases with city size. Mansueto Institute for Urban Innovation Research Paper Forthcoming. 2020.
23. Carozzi F. Urban density and COVID-19. 2020.
24. Chu DK, Akl EA, Duda S, Solo K, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020.
25. Gandhi M, Beyrer C, Goosby E. Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer. Journal of general internal medicine. 2020;35(10):3063-6.
26. Ma QX, Shan H, Zhang L, et al. Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2.
27. Grimm P. Social desirability bias. Wiley international encyclopedia of marketing. 2010.
28. Nisar MI, Ansari N, Amin M, et al. Serial population based serosurvey of antibodies to SARS-CoV-2 in a low and high transmission area of Karachi, Pakistan. medRxiv. 2020.
29. Dawn News. 300,000 persons infected with Covid-19 in Islamabad: survey 2020 [cited 2020 22 December]. Available from: 300,000 persons infected with Covid-19 in Islamabad: survey
30. Pakistan Go. COVID-19 Tracker (Press Release, July 2020) 2020 [cited 2020 22 December]. Available from: http://covid.gov.pk/press-release-details/2.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	3, 4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	3, 4
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	3, 4
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4
Bias	9	Describe any efforts to address potential sources of bias	4
Study size	10	Explain how the study size was arrived at	3, 4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7, 8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7, 8
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	7
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of 0cases and controls was addressed	NA

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	4
		(b) Give reasons for non-participation at each stage	4
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	4
		(b) Indicate number of participants with missing data for each variable of interest	5, 6
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	4-7
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-7
		(*b*) Report category boundaries when continuous variables were categorized	5-7 (Tables 2 & 3)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	12

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055381.R1
Article Type:	Original research
Date Submitted by the Author:	19-Jan-2022
Complete List of Authors:	Ahmad, Ahsan ; Center for Global Public Health- Pakistan, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan Shahzad, Khurram; Health Services Academy Masood, Mariumn; Health Services Academy Umar, Maida; Health Services Academy Abbasi, Fahad; Health Services Academy Hafeez, Assad; Health Services Academy
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Immunology (including allergy)
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, Epidemiology < TROPICAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study

Corresponding Author: Dr. Mariumn Masood, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan. Email: msmasood148@gmail.com

Author Information

1. Dr. Ahsan M Ahmed, Center for Global Public Health- Pakistan, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
ahsen@cgph.org.pk

2. Dr. Khurram Shahzad, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
baig.khurram@gmail.com

3. Dr. Mariumn Masood, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
msmasood148@gmail.com

4. Ms. Maida Umar, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
maidakhawja@gmail.com

5. Dr. Fahad Abbasi, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
idrfahadabbasi@gmail.com

6. Dr. Assad Hafeez, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
az10@hotmail.com

Word Count: 3414 words

Ethics Approval: from the Institutional Ethical Review Committee of Health Services Academy, Islamabad, Pakistan (Location: Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan)

ABSTRACT

Objectives This study adapted The World Health Organization's 'Unity Study' protocol to estimate the population prevalence of antibodies to Severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) and risk factors for developing SARS-CoV-2 infection.
Design This population-based, age-stratified cross-sectional study was conducted in households (HH).
Participants 4,998 households out of 6,599 consented (one individual per household). 51% were male participants. All ages and sexes were eligible. (Exclusion Criteria: contraindications to venipuncture. However no such case was encountered).
Primary and Secondary Outcome Measures
Following were the Planned Outcome Measures
Primary Indicators: i) Seroprevalence (population and age-specific); ii) Asymptomatic fraction (proportion of cases that are asymptomatic)
Secondary Indicators**:** i) Population groups most at risk for SARS-CoV-2-infection
Following were the Measured Outcome Measures. These were different from the planned indicators (i.e. two out of the three planned indicators were measured) due to operational reasons and time constraints.
Primary Indicators: i) Seroprevalence (population and age-specific)
Secondary Indicators: Population groups most at risk for SARS-CoV-2-infection
Results Overall seroprevalence of SARS-CoV-2 antibodies was 7.1%. 6.3% of individuals were IgG positive while IgM positivity was 1.9%. The seroprevalence among different age groups ranged from 3.9% (0-9 years) to 10.1% (40-59 years). Seroprevalence in districts ranged from 0% (Ghotki) to 17% (Gilgit). History of contact with a confirmed COVID-19 case, urban-residence and mask-use were key risk factors for developing SARS-CoV-2 infection.
Conclusions This survey provides useful estimates for seroprevalence in the general population and information on risk factors for developing SARS-CoV-2 infection in the country. It is premised that similar studies need to be replicated at the population level on a regular basis to monitor the disease and immunity patterns related to COVID-19.

Strengths and Limitations of the Study
Strengths:

- A large sample size of about 5000 individuals was taken with random selection at each stage i.e. district, Union Council, Villages (for rural Union Councils) households, and individuals, and representation from all four provinces and both regions of the country
- There is an almost equal representation from both the sexes; and the rural and urban samples were in accordance with the proportion of urban: rural Union Councils of each district.

Limitations

- Seroprevalence was gauged using Rapid Diagnostic Tests (RDTs), which are not the gold standard (for this purpose) and may have varying sensitivity depending on the time since infection onset

- Measurement of some of the risk factors for developing SARS-CoV-2 infection was through self-report, which have the potential to introduce recall bias and social desirability bias.
- The age structure of the study population differs from the demographic distribution of the population in the country, which may have caused

INTRODUCTION

The Coronavirus disease 2019 (COVID-19), originating from Wuhan, China was declared a pandemic by the World Health Organization (WHO) on the 11th of March, 2020.(1)] Pakistan reported its first case on February 26th, 2020, witnessing its first peak during July, 2020.(2)] Among various response measures in Pakistan have been the conduction of research studies to inform response measures to COVID-19 and to enable a better understanding of its epidemiology and spread. This includes partaking in the WHO's global research initiatives such as the 'WHO Unity Studies' through the conduction of national seroprevalence studies.(3)]These study protocols have also been adapted around the world at various geographic levels to provide contextual data on the evolving pandemic.

Numerous seroprevalence studies have been conducted around the globe since the onset of the pandemic, with more evidence from large-scale nationwide studies being reported as the situation evolves, and the results showing wide contextual variations.(4)]In neighboring countries, Iran reported a high seroprevalence of 17.1% in a large study conducted across 17 provinces during April to June 2020, although the results were based on, and were from a much earlier phase of the pandemic.(5) India also conducted a national seroprevalence survey (for IgG) in adults during the same period utilizing ELISA (enzyme-linked immunosorbent assay). However, in stark contrast to Iran it reported markedly lower seroprevalence of 0.7% in its study population.(6)] Studies in the other regions also depict varying patterns of seroprevalence depending on the timelines of the pandemic in their respective countries.(7-9)]One of the largest seroprevalence studies around the globe in England has reported a decline in seroprevalence estimates of IgG by 26.5% between June and September 2020, from 6.0% to 4.4% by lateral flow immunoassay (LFIA) tests.(10)]

This study in Pakistan is part of the global 'WHO Unity Studies' initiative with the main objective of estimating the seroprevalence of IgG and IgM antibodies to the 'SARS-CoV-2' coronavirus in the general population in the country. This study will not only provide data regarding the exposure of the general population to COVID-19, but would also shed light on some risk factors for developing SARS-CoV-2 infection. Overall, we expect that the estimates would provide us with ample evidence to gauge the population-level scenario of COVID-19 in the country as well as provide insights into other epidemiological aspects of the disease, including the risk factors for developing SARS-CoV-2 infection.

METHODS

Patient and Public involvement Statement: Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

This population-based, age-stratified cross-sectional study, was conducted at the level of households (HH) between October 21st and November 8th, 2020, with Ethics Approval from the Institutional Ethical Review Committee of Health Services Academy, Islamabad.

This was a nationwide study in the four provinces (Punjab, Sindh, Baluchistan and Khyber Pakhtunkhwa/KP) and two regions (Azad Jammu & Kashmir (AJK) and Gilgit Baltistan (GB)) of Pakistan. Individuals of all ages and genders were eligible to participate.

Sample size calculation was done using Open Source Epidemiologic Statistics for Public Health [a] using multistage sampling design (with expected prevalence of 11% which were preliminary estimates from a previous national level study;(11)]) difference between upper and lower limit of the interval estimate as 1.25% (0.75% on either side); and a design effect of '2'). The resultant sample size of 4,803 was rounded off to 5,000 and distributed equally among the ten districts under study as well as among all age brackets.

The study was conducted in ten districts of the country according to the following criteria: one high and one low prevalence district each was selected from provinces; and the highest-prevalence district each was selected from regions AJK and GB. In Pakistan's context, high prevalence was taken as a cumulative of more than 500 COVID-19 cases (Tier 1); and low prevalence as a cumulative of less than 500 cases (Tier 2) by the cut-off date of June 30th, 2020. The population of district and other characteristics were not used during sampling due to variable testing rates by population in various districts.

After district selection, Union Council (UC) selection was done with the aim of recruiting 500 households/participants randomly from each district (25 participants defined as one cluster). UCs were randomly selected from each district; the selection was in accordance with the "Urban UC: Rural UC" ratio for each district. Where there were less than 20 UCs in a district, the number of clusters were increased to make the total equal to 500 participants per district. Systematic random sampling was employed for the next stage i.e. household selection. Thereafter, one individual was randomly selected from each consenting household in line with the age distribution of the study. Each cluster of 25 had five participants from each age group. In this way 25 HH/individuals were recruited for each cluster, with random selection at different stages (as described) aiming to reduce potential sources of selection bias.

Data collection for the survey included on-the-spot recruitment of households in the selected localities. After obtaining informed consent, each participant(one individual randomly selected out of all the eligible from each household) was asked to provide information to the enumerator to fill a pre-tested questionnaire. The questionnaire was adopted from WHO Unity Studies protocol and collected information of socio-demographic variables, medical and symptom history, preventive behaviours (note: handwashing was defined as follows- individuals were inquired about number of times they had washed their hands with soap for 20 seconds; the variable was categorized into those washing hands at least six times and those less than six times.), complications and history of recent death in family followed by Rapid Diagnostic Test (RDT) results performed by a trained

[a] https://www.openepi.com/Menu/OE_Menu.htm

phlebotomist. (3)] RDT was performed with 'Bioperfectus' kits for IgG/IgM, and the results were provided to each participant on the spot(within a few minutes). Prior to the survey, data collectors and field teams underwent training, which included PPE usage for infection prevention and control. Monitoring visits were also conducted during the data collection period to enhance data quality and results. We computed descriptive statistics and univariate logistic regression that examined associations of seropositivity with age, gender, location, tier, gender and symptoms with 95% confidence intervals (CI).

The multivariate binary logistic regression model

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \qquad (1)$$

where p is the probability that an individual is seropositive, β_0 is the intercept, β_i are the coefficients and X_i represents the independent variables. The outcome variable represents the seropositivity which is a binary categorical variable whereas location, age, mask use and contact with COVID-19 positive person are the predictor variables. A forward-stepwise process was utilized for the selection of significant variables in the final model. During the analysis, some interaction terms were considered but not included in the final model because they were not statistically significant. The selection of multivariate logistic model was based on Homer-Lemeshow goodness of fit, biological interpretability and statistical significance. The significance level or alpha was 0.05. Data analysis for the survey was performed using SPSS version 23.

RESULTS

During field work, recruitment was continued till the target sample was achieved. A total of 6,599 households were reached for on-the-spot participation, of which 1,601 (24%) did not provide consent to participate. A total of 4,998 households (with one individual per household) consented to participate in the study across the ten selected districts of Pakistan. Individuals of all ages and genders were eligible. The proportion of males in the recruited participants was 51%. The mean age of males (31.7 years) was similar to that of females (32.8 years). Almost two thirds (62%) of individuals were recruited from rural areas. Male to female ratio of sampled individuals was similar for rural and urban areas.

Information on seroprevalence was available for all 4998 participants with no missing values. The overall seroprevalence of SARS-CoV-2 antibodies was 7.1%. Almost 6.3% of individuals were IgG positive while IgM positivity was 1.9%. Seroprevalence in districts ranged from 0.0% (Ghotki) to 17.0% (Gilgit). Most of the districts reported a range of 6% to 9% (**Error! Reference source not found.**).The seroprevalence among different age groups ranged from 4% (0-9 years) to 10% (40-59 years). (**Error! Reference source not found.**). A total of 4% reported to have had contact with a COVID-19 positive individual.

Table 1 : Seroprevalence at District Level

District	Reported prevalence based on RT-PCR[a]	Positive (%)	Negative (%)
Quetta	High	16(3.2)	484(96.8)
Mardan		31(6.2)	470(93.8)
Rawalpindi		34(6.8)	467(93.2)
Ghotki		0 (0.0)	498(100)
Muzaffarabad	Low	39(8.6)	460(92.2)
Gilgit		85(17.0)	415(83.0)
Sibbi		24(4.8)	474(95.2)
Abbottabad		42(8.4)	460(91.6)
Lodhran		43(8.6)	457(91.4)
Jacobabad		37(7.4)	462(92.6)

Table 2 Seroprevalence by age, place, district tier, gender and (significant)symptoms

Variables		Positive (%)	Negative (%)	Odds ratio (95% CI)	p-value
Seroprevalence for COVID-19		351(7.1)	4647(93.0)		
Gender (*n*=4,998)	Male	178(6.9)	2394(93.1)	0.97(0.78-1.20)	0.879
	Female	173(7.1)	2253(92.9)		
Age (years) (*n*=4,997)	0 – 9	38(3.9)	935(96.1)		
	10 – 19	46(4.5)	973(95.5)	0.43(0.29-0.63)	0.000
	20 – 39	80(7.8)	950(92.2)	0.49(0.34-0.71)	0.000
	40 – 59	101(10.1)	899(89.9)	0.88(0.64-1.21)	0.390
	60+	85(8.7)	890(91.3)	1.18(0.87-1.59)	0.309
Location (*n*=4,998)	Urban	154(8.1)	1741(91.9)	1.30(1.04-1.62)	0.014
	Rural	197(6.3)	2906(93.7)		
District prevalence (reported cases) (*n*=4,998)	Low prevalence (tier-2)	270(9.0)	1919(96.0)	2.34(1.81-3.02)	0.000
	High prevalence (tier-1)	81(4.1)	2728(91.0)		
Contact with COVID-19 positive case (*n*=4,971)	Yes	25(13.9)	155(86.1)	2.34(1.81-3.03)	0.001
	No	292(6.7)	4041(93.3)		
Sore throat (*n*=4,987)	Yes	118(10.6)	998(89.4)	1.86(1.46-2.32)	0.000
	No	233(6.0)	3638(94)		
Fatigue (*n*=4,991)	Yes	43(10.9)	351(89.1)	1.706(1.22-	0.002

	No	308(6.7)	4289(93.3)	2.39)	
Joint ache (*n*=4,988)	Yes	65(10.7)	541(89.3)	1.734(1.31-2.30)	0.000
	No	284(6.5)	4.98(93.5)		
High grade fever (*n*=4,986)	Yes	123(8.6)	1313(91.4)	1.365(1.09-1.72)	0.008
	No	228(6.4)	3322(93.6)		
Cough (*n*=4,993)	Yes	98(8.5)	1055(91.5)	1.317(1.03-1.67)	0.026
	No	253(6.6)	3587(93.4)		
Runny nose (*n*=4,987)	Yes	97(8.5)	1044(91.5)	1.32(1.03-1.7)	0.026
	No	253(6.6)	3593(93.4)		

The use of preventative behaviors was also studied. The use of face-masks while going out in public was reported to be 63%. Mask use was similar in urban and rural areas (63%). It was highest in 20-59 years age group (68%), while 60+ group reported relatively less use (49%). Mask use increased incrementally with education, from 36% in those non-educated to 82% in individuals above matric (ten years of education). Handwashing (washing hands at least six times with soap and water for 20 seconds in last 24 hours) was reported relatively less compared to mask use (39%). Handwashing was higher in urban (44%) areas; among females (43%); and increased with education, being highest in individuals above matric (**Error! Reference source not found.**).

Table 3 Mask use and handwashing practices by socio-demographic characteristics

Variables		**Mask use (%)**			**Handwashing (%)[a]**		
		Yes	**No**	**Total(%)**	**Yes**	**No**	**Total(%)**
Overall		3,128(62.6)	1,844(37.1)		1,946(38.9)	3,052(61.1)	
Location	Urban	1,185(62.9)	699(37.1)	1,884(37.8)	839(44.3)	1,056(55.7)	1,895(37.9)
	Rural	1,943(62.9)	1,145(37.1)	3,088(62.1)	1,107(35.7)	1,996(64.3)	3,103(62.1)
District Tiers	Tier-1	1,121(56.5)	864(43.5)	1,985(39.9)	762(38.1)	1,238(61.9)	2,000(40.0)
	Tier-2	2,007(67.2)	980(32.8)	2,987(60.0)	1,184(39.5)	1,814(60.5)	2,992(60)
Gender	Male	1,654(64.6)	907(35.4)	2,561(51.5)	905(35.2)	1,667(33.4)	2,572(51.5)
	Female	1,474(61.1)	937(38.9)	2,411(48.5)	1,041(42.9)	1,385(27.7)	2,426(48.5)
Age	0-19	576(59.5)	392(40.5)	968(24.5)	264(27.1)	709(72.9)	973(24.5)
	20-59	1,364(67.5)	657(32.5)	2,021(51.5)	934(46.0)	1,096(54.0)	2,030(51.0)
	60+	476(49.1)	50.9(50.9)	970(24.5)	382(39.2)	593(2398)	975(24.5)
Education	No education	412(36.0)	731(39.0)	1,143(36.9)	326(28.4)	822(71.6)	1,148(36.9)
	Primary	194(61.0)	124(24.8)	318(10.3)	137(42.9)	182(57.1)	319(10.3)
	Matric	491(75.2)	162(18.4)	653(21.2)	307(46.5)	353(53.5)	660(21.2)

	Above matric	797(81.6)	180(64.0)	977(36.9)	573(58.4)	408(13.1)	981(31.6)

[a]

Symptoms during past three months were inquired to look for possible association with COVID-19 seropositivity. Total of 23 symptoms were inquired relating to multiple systems. The symptoms shown in **Error! Reference source not found.** were significantly higher in seropositive individuals. Sore throat, fatigue, and joint aches were strongly associated with seropositivity. Among COVID-19 seropositive individuals, 68% had at least one symptom in last two months, while 32% reported to be completely asymptomatic during this period.

Almost 24% of individuals reported having at least one comorbidity. Hypertension was reported the most (18%), followed by diabetes (5%) and chronic kidney disease (2%). The reported occurrence of heart disease (1%) and asthma (2%) was relatively lower. Reported prevalence of at least one comorbidity increased with age with maximum being reported for 60+ age group (54%), followed by 40-59 years age group (40%).

Multivariate logistic regression analysis was performed to identify factors associated with seropositivity (Table 4). Urban residents were more likely to test positive for COVID-19 antibodies than rural residents (OR 1.29, 95% CI 1.04 - 1.61). Individuals aged 20 and above were about twice as likely to be seropositive than those who were 0-9 years old. Odds of seropositivity were also high among individuals who did not wear face mask (OR 1.54, 95% CI 1.20 – 1.975) and in those who reported contact with COVID-19 person (OR 1.81, 95% CI 1.16 – 2.83).

For any discrepancy and to ensure completeness of data, field teams were contacted and cross-checked. All cases are included in the analysis and missing data upto 1% were considered acceptable because of low occurrence.

Table 4 Factors associated with seropositivity

Variables		**b**	**S.E.**	**Adjusted OR (95% CI)[a]**	**p-value**
Location	Urban	0.256	0.113	1.29(1.035-1.612)	0.024
	Rural[b]				
Age (years)	0 – 9[b]				
	10 – 19	0.112	0.225	1.11(0.7-1.737)	0.618
	20 – 39	0.650	0.204	1.92(1.285-2.854)	0.004
	40 – 59	0.975	0.197	2.65(1.803-3.899)	0.000
	60+	0.894	0.202	2.45(1.646-3.630)	0.000
Mask use	Yes	0.434	0.126	1.54(1.205-1.975)	0.001
	No[b]				
Contact with COVID-19 positive case	Yes	0.596	0.227	1.81(1.163-2.831)	0.009
	No [b]				

Constant		-3.599			

[a] Hosmer-Lemshow goodness-of-fit test: chi-square value = 8.322, P = 0.403

[b] Reference category

DISCUSSION

The survey through the use of Unity Studies' age-stratified approach, estimated the national seroprevalence of IgG and IgM antibodies for COVID-19 to be 7.1% in Pakistan based on RDT testing. Among the included districts, the highest prevalence was observed for Gilgit, followed by Lodhran and Muzaffarabad (Table 1). This study was initiated in the last week of October 2020 and field activities were completed by second week of November 2020. This was a time frame, when the first wave of the pandemic was considered to have largely subsided, the new number of cases per day was markedly lower, and there was a threat of a second wave of the pandemic in the forthcoming winter months of November onwards. During this time frame < 1000 confirmed cases per day were being recorded, in the backdrop of the highest daily number of cases (6,825) that had been reported on 13th June, 2020.(2)]

Although other tests such as ELISA (Enzyme-linked immunosorbent assay) offer greater accuracy than RDTs in terms of antibody detection,(12)] due to practical and operational issues, RDTs were opted for this large-scale population-based study to estimate the seroprevalence of antibodies to COVID-19, with the results for on-the-spot testing in the field available within twenty minutes. Some researchers using similar kits to those used in this survey reported the sensitivity (to detect IgG/IgM)to range between 41% (at 1-5 days since symptom onset in patients positive by RT-PCR) to 100% (at > 20 days since symptom onset in patients positive by RT-PCR); while the reported specificity when compared with PCR was 95%(13) ()] This time-dependent sensitivity of RDTs to detect SARS-Cov-2 antibodies has also been noted elsewhere.(14)] Due to the widely varying sensitivity of the testing method, it is likely that the actual seroprevalence may have been much higher than this study's estimates.

It should be noted that the districts' population size and the age structure of the districts' population were not taken into account during district selection or selection of individuals of different age groups from within households, this may have potentially introduced bias in the seroprevalence estimates.

Additionally, 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 limitations.

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

About one third of seropositive individuals had reported to have experienced symptoms during the past two months. Six out of the studied 23 symptoms experienced during the past two months were found to be significantly associated with seropositivity in the univariate analysis, most of which were respiratory/pharyngeal symptoms including sore throat, shortness of breath, cough and runny nose. High grade fever, joint aches and fatigue were the three generalized symptoms associated with seropositivity. Although the symptoms were self-reported with a possibility for recall bias, information on symptoms was obtained before the testing was done hence it is likely that any misclassification may have been non-differential.

Similar to what has been reported in some other national studies (including a large-scale national-level household study by Pollán et al in Spain), gender was not found to be significantly associated with seropositivity for COVID-19 antibodies.(8, 9, 15)] The association of age with seropositivity increased with age until 59 years, and declined slightly in those above this age bracket. As observed previously, nasal gene expression of angiotensin-converting enzyme 2 (ACE2) has been postulated to be responsible for this age-related pattern, and children have been reported to be less susceptible to contracting COVID-19 than adults potentially due to the role of innate immunity(16); and the protective effect of lower levels of this enzyme in their nasal epithelium, as also reported elsewhere.(17, 18)]

The risk of seropositivity doubled in those with a history of exposure to a diagnosed COVID-19 patient and was found to be statistically significant. A study in Italy reported an even higher Odds ratio of 2.5 in those who had previous contact with a case. (19)] Association of COVID-19 with household contact with known cases of COVID-19 has also found to be significant in other prevalence studies.(20)].

Differences in seroprevalence between Tier 1 (districts considered to be high transmission areas by the end of June 2020) and Tier-2 districts (districts considered to be low transmission areas at the end of June 2020), were significant at about 4.9%. However, it is important to note that the seroprevalence was lower in Tier 1 districts in this study. This pattern suggests that the transmission scenarios would have evolved in three-four months since the chosen cut-off date, and the areas earlier considered to be higher transmission at the end of June (i.e. Tier 1 districts) may have now become areas of relatively lower transmission, and vice versa. The said difference is understood to be possibly due to a higher proportion of population in previously high risk districts, having possibly experienced the exposures, with possibility of reduction in IgM levels across a time span of 90 or more days (i.e. the time span between June and October 2020).

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 2.9 million 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 kilometer)

Urban and rural areas were sampled from each district in accordance with the urban to rural ratio for that particular district. Overall, urban areas reported a higher prevalence (8.1%) than rural areas (6.3%), and urban residents were more likely to be seropositive for COVID-19 than rural residents (Table 4). While some researchers have argued that larger city sizes tend to have higher attack rates, (21)] other studies have gone further so as to report that while urban areas do have a propensity for earlier outbreaks than rural areas, population density is not significantly associated with COVID-19 cases.(22)] Our results seem to align with the former, i.e. showing a significant association of urban residence with COVID-19 seropositivity.

Among studied behaviors, use of masks was reported by about two-thirds of the study participants with similar values in rural and urban areas. Mask use was found to be significantly linked with seropositivity in logistic regression in our study (Table 4), in line with wide-ranging evidence.(1, 23, 24) On the other hand, hand-washing(at least six times per day) had a lower prevalence (39%) in the study population and was not found to be a significant risk factor in this study, although other researchers have reported protective benefits of hand hygiene (25)]. Since these have both been recommended preventive behaviours during the pandemic, the results may have been affected by Social Desirability Bias(26)] in the study population, causing over-reporting of these behaviours. In such a case, any true association of these behaviours with seropositivity for COVID-19 may have been masked, if they do indeed influence seropositivity for SARS-CoV-2.

Various studies to estimate the seroprevalence of COVID-19 have been conducted in the country at different scales. For example, preliminary results from a national sero-prevalence study that was conducted during July 2020 reflected a prevalence of 11.2%.(11)] Seroprevalence studies at a smaller scale in the country have also been conducted in Karachi city and Islamabad Capital Territory(ICT), with the former reporting figures of 9.7% and 15.1% in 'low-transmission' and 'high-transmission' areas in the city, respectively,(27)] and the latter reporting an overall seroprevalence of 14.5%.(28, 29)] Results from the aforementioned sub-national surveys differ from those of the current study in which the seroprevalence was found to 4.1% in Tier-1 districts and 9.2% in Tier-2 districts, respectively. This may have been due the differing time frames of the studies in Karachi and ICT, which were conducted during the first wave of COVID-19 in the country; while the current study was conducted during October-November 2020, when the first wave of the pandemic had largely subsided.

Overall, it is likely that the seroprevalence estimates may have been affected by low sensitivity of the testing methods(causing underestimation of the values); and the estimates for association of seropositivity with risk factors for developing SARS-CoV-2 infection may have been affected by biases including recall and social desirability bias among the study population. The seroprevalence estimates provided by this study may be interpreted with caution as an estimated value for the general population, particularly since the age structure of the study population differs from the demographic distribution of the population in the country.

CONCLUSIONS

Notwithstanding the limitations of the study, this survey provides useful prevalence estimates as well as information on risk factors for developing SARS-CoV-2 infection. The results also show that the youngest age groups have the lowest proportion of sero-positivity as compared to those aged 40 years and above. Interestingly, Tier 1 districts (considered to be high risk based on the number of PCR test based confirmed cases by the end of June 2020) reflected lower prevalence as compared to Tier 2 districts which may perhaps be depicting reversing patterns at the population level. A history of contact with a confirmed COVID-19 case, being an urban resident and mask use were key risk factors for developing SARS-CoV-2 infection . Keeping view of these findings, it is premised that similar studies need to be replicated at the population level on a regular basis to monitor the disease and immunity patterns related to COVID-19.

List of Abbreviations

COVID-19- The Coronavirus disease of 2019

WHO- World Health Organization

ELISA - Enzyme-linked immunosorbent assay

RDT - Rapid Diagnostic Test

SARS-CoV-2 - Severe acute respiratory syndrome coronavirus 2

RT-PCR - Reverse transcription polymerase chain reaction

ACE2- Angiotensin-converting enzyme 2

HH- Household

OR- Odds Ratio

IRB - Institutional Review Board

KP - Khyber Pakhtunkhwa

AJK- Azad Jammu & Kashmir

GB- Gilgit Baltistan

UC- Union Council

LFIA - Lateral flow immunoassay

- **Competing interests**

  The authors declare that they have no competing interests

- **Funding**

This study was funded by Health Services Academy (HSA), Islamabad which is affiliated with the Ministry of National Health Services Regulations and Coordination, Government of Pakistan. HSA was involved in all aspects of the survey including study design/technical supervision; collection, analysis and interpretation of the data; in the writing of the manuscript; and in the decision to submit the paper for publication.

Funding/grant award number: F-11-2020-WHO/DAi/HSA

- **Authors' contributions**

AH and AA supervised the overall process; AA contributed to the analysis, manuscript writing and monitoring of data collection; KS led the coordination of all survey activities including design, training, administrative aspects and data analysis; led the process of preparing tools for data collection; contributed to training of data collectors and monitoring of the data collection process, and to the analysis and reporting of Results; MM supported preparation of data collection tools, contributed to training of data collectors, drafted the Household sampling methodology, conducted household refusal data analysis, contributed to monitoring of data collection, conducted literature reviews to analyse epidemiological data from the survey, and prepared the initial manuscript draft; MU contributed to data management, analysis and reporting. FA supported preparation of data collection tools, contributed to monitoring of data collection and training of data collectors. All authors reviewed and approved the manuscript.

- **Acknowledgements**

Not Applicable

- **Data Sharing Statement**
Data are available upon reasonable request

These are de-identified participant data, available from Health Services Academy(HSA) (Dr. Shahzad Ali Khan, Vice Chancellor, HSA. Email: shahzad@hsa.edu.pk). Data may be reused for future meta-analysis upon reasonable justification of benefit at regional/global level after permission. Additional information is also available (protocol inclusive of study tool, and statistical analysis plan).

- **Ethics**
This study involves human participants and was approved by an Ethics Committee(s) or Institutional Board(s)

Please state the name of the Ethics Committee(s) or Institutional Board(s) that approved this study with reference number or ID for your ethics approval OR the reason one was not provided:

Institutional Ethical Review Committee of Health Services Academy, Ministry of National Health Services, Islamabad, Pakistan

No. 7-82/IERC-HSA/2020-33

References

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 2020 [cited 2020 22 December]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
2. Government of Pakistan. COVID-19 Dashboard 2020 [cited 2020 22 December]. Available from: http://covid.gov.pk/stats/pakistan?locale=en.
3. World Health Organization. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigation Protocols 2020 [cited 2020 22 December]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations.
4. Arora RK, Joseph A, Van Wyk J, Rocco S, Atmaja A, May E, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. The Lancet Infectious Diseases. 2020.
5. Poustchi H, Darvishian M, Mohammadi Z, Shayanrad A, Delavari A, Bahadorimonfared A, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. The Lancet Infectious Diseases. 2020.
6. Murhekar MV, Bhatnagar T, Selvaraju S, Rade K, Saravanakumar V, Thangaraj JWV, et al. Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020. Indian Journal of Medical Research. 2020;152(1):48.
7. Xu X, Sun J, Nie S, Li H, Kong Y, Liang M, et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. Nature Medicine. 2020:1-3.
8. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. The Lancet. 2020;396(10250):535-44.
9. Herzog S, De Bie J, Abrams S, Wouters I, Ekinci E, Patteet L, et al. Seroprevalence of IgG antibodies against SARS coronavirus 2 in Belgium: a prospective cross-sectional study of residual samples. medRxiv. 2020.
10. Ward H, Cooke G, Atchison CJ, Whitaker M, Elliott J, Moshe M, et al. Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. MedRxiv. 2020.
11. Dawn News. 11pc Pakistanis have developed protective immunity: study 2020 [cited 2020 22 December]. Available from: https://www.dawn.com/news/1575607/11pc-pakistanis-have-developed-protective-immunity-study.
12. Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. Diagnostics. 2020;10(5):319.
13. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, et al. Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. Nature Biotechnology. 2020;38(10):1174-83.
14. Rashid ZZ, Othman SN, Samat MNA, Ali UK, Wong KK. Diagnostic performance of COVID-19 serology assays. The Malaysian Journal of Pathology. 2020;42(1):13-21.
15. Lai C-C, Wang J-H, Hsueh P-R. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: An up-to-date review. International Journal of Infectious Diseases. 2020.
16. Yang HS, Costa V, Racine-Brzostek SE, Acker KP, Yee J, Chen Z, et al. Association of age with SARS-CoV-2 antibody response. JAMA network open. 2021;4(3):e214302-e.
17. Patel AB, Verma A. Nasal ACE2 levels and COVID-19 in children. Jama. 2020;323(23):2386-7.
18. Bunyavanich S, Do A, Vicencio A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. Jama. 2020.

19. Vena A, Berruti M, Adessi A, Blumetti P, Brignole M, Colognato R, et al. Prevalence of antibodies to SARS-CoV-2 in Italian adults and associated risk factors. Journal of clinical medicine. 2020;9(9):2780.
20. Menachemi N, Yiannoutsos CT, Dixon BE, Duszynski TJ, Fadel WF, Wools-Kaloustian KK, et al. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. Morbidity and Mortality Weekly Report. 2020;69(29):960.
21. Stier A, Berman M, Bettencourt L. COVID-19 attack rate increases with city size. Mansueto Institute for Urban Innovation Research Paper Forthcoming. 2020.
22. Carozzi F. Urban density and COVID-19. 2020.
23. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020.
24. Gandhi M, Beyrer C, Goosby E. Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer. Journal of general internal medicine. 2020;35(10):3063-6.
25. Ma QX, Shan H, Zhang L, Li M, Yang M, Chen JM. Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2.
26. Grimm P. Social desirability bias. Wiley international encyclopedia of marketing. 2010.
27. Nisar MI, Ansari N, Amin M, Khalid F, Hotwani A, Rehman N, et al. Serial population based serosurvey of antibodies to SARS-CoV-2 in a low and high transmission area of Karachi, Pakistan. medRxiv. 2020.
28. Dawn News. 300,000 persons infected with Covid-19 in Islamabad: survey 2020 [cited 2020 22 December]. Available from: 300,000 persons infected with Covid-19 in Islamabad: survey
29. Pakistan Go. COVID-19 Tracker (Press Release, July 2020) 2020 [cited 2020 22 December]. Available from: http://covid.gov.pk/press-release-details/2.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4, 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4, 5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4, 5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4, 5
Bias	9	Describe any efforts to address potential sources of bias	4, 5
Study size	10	Explain how the study size was arrived at	3, 4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5- 8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7, 8
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	8
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of 0cases and controls was addressed	NA

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	5, 6
		(b) Indicate number of participants with missing data for each variable of interest	5, 6
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	5-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-8
		(*b*) Report category boundaries when continuous variables were categorized	6-7 (Tables 2 & 3)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-055381.R2
Article Type:	Original research
Date Submitted by the Author:	24-Feb-2022
Complete List of Authors:	Ahmad, Ahsan ; Center for Global Public Health- Pakistan, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan Shahzad, Khurram; Health Services Academy Masood, Mariumn; Health Services Academy Umar, Maida; Health Services Academy Abbasi, Fahad; Health Services Academy Hafeez, Assad; Health Services Academy
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health, Immunology (including allergy)
Keywords:	COVID-19, Public health < INFECTIOUS DISEASES, Epidemiology < TROPICAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

COVID-19 Seroprevalence in Pakistan: a Cross-Sectional Study

Corresponding Author: Dr. Mariumn Masood, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan. Email: msmasood148@gmail.com

Author Information

1. Dr. Ahsan M Ahmed, Center for Global Public Health- Pakistan, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
ahsen@cgph.org.pk

2. Dr. Khurram Shahzad, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
baig.khurram@gmail.com

3. Dr. Mariumn Masood, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
msmasood148@gmail.com

4. Ms. Maida Umar, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
maidakhawja@gmail.com

5. Dr. Fahad Abbasi, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
idrfahadabbasi@gmail.com

6. Dr. Assad Hafeez, Health Services Academy, Ministry of National Health Services Regulations and Coordination, Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan
az10@hotmail.com

Word Count: 3414 words

Ethics Approval: from the Institutional Ethical Review Committee of Health Services Academy, Islamabad, Pakistan (Location: Opposite National Institute of Health (NIH) Chak Shahzad, Islamabad, Pakistan)

ABSTRACT

Objectives This study adapted The World Health Organization's 'Unity Study' protocol to estimate the population prevalence of antibodies to Severe acute respiratory syndrome coronavirus 2 (SARS CoV-2) and risk factors for developing SARS-CoV-2 infection.
Design This population-based, age-stratified cross-sectional study was conducted in households (HH).
Participants 4,998 households out of 6,599 consented (one individual per household). 51% were male participants. All ages and sexes were eligible. (Exclusion Criteria: contraindications to venipuncture. However no such case was encountered).
Primary and Secondary Outcome Measures
Following were the Measured Outcome Measures. These were different from the planned indicators (i.e. two out of the three planned indicators were measured) due to operational reasons and time constraints.
Primary Indicators: i) Seroprevalence (population and age-specific)
Secondary Indicators: Population groups most at risk for SARS-CoV-2-infection
Results Overall seroprevalence of SARS-CoV-2 antibodies was 7.1%. 6.3% of individuals were IgG positive while IgM positivity was 1.9%. The seroprevalence among different age groups ranged from 3.9% (0-9 years) to 10.1% (40-59 years). Seroprevalence in districts ranged from 0% (Ghotki) to 17% (Gilgit). History of contact with a confirmed COVID-19 case, urban-residence and mask-use were key risk factors for developing SARS-CoV-2 infection.
Conclusions This survey provides useful estimates for seroprevalence in the general population and information on risk factors for developing SARS-CoV-2 infection in the country. It is premised that similar studies need to be replicated at the population level on a regular basis to monitor the disease and immunity patterns related to COVID-19.

Strengths and Limitations of the Study
Strengths:

- A large sample size of about 5000 individuals was taken with random selection at each stage i.e. district, Union Council, Villages (for rural Union Councils) households, and individuals, and representation from all four provinces and both regions of the country
- There is an almost equal representation from both the sexes; and the rural and urban samples were in accordance with the proportion of urban: rural Union Councils of each district.

Limitations

- Seroprevalence was gauged using Rapid Diagnostic Tests (RDTs), which are not the gold standard (for this purpose) and may have varying sensitivity depending on the time since infection onset
- Measurement of some of the risk factors for developing SARS-CoV-2 infection was through self-report, which have the potential to introduce recall bias and social desirability bias.
- The age structure of the study population differs from the demographic distribution of the population in the country, which may have caused

INTRODUCTION

The Coronavirus disease 2019 (COVID-19), originating from Wuhan, China was declared a pandemic by the World Health Organization (WHO) on the 11th of March, 2020.(1) Pakistan reported its first case on February 26th, 2020, witnessing its first peak during July, 2020.(2) Among various response measures in Pakistan have been the conduction of research studies to inform response measures to COVID-19 and to enable a better understanding of its epidemiology and spread. This includes partaking in the WHO's global research initiatives such as the 'WHO Unity Studies' through the conduction of national seroprevalence studies.(3) These study protocols have also been adapted around the world at various geographic levels to provide contextual data on the evolving pandemic.

Numerous seroprevalence studies have been conducted around the globe since the onset of the pandemic, with more evidence from large-scale nationwide studies being reported as the situation evolves, and the results showing wide contextual variations.(4) In neighboring countries, Iran reported a high seroprevalence of 17.1% in a large study conducted across 17 provinces during April to June 2020, although the results were based on, and were from a much earlier phase of the pandemic.(5) India also conducted a national seroprevalence survey (for IgG) in adults during the same period utilizing ELISA (enzyme-linked immunosorbent assay). However, in stark contrast to Iran it reported markedly lower seroprevalence of 0.7% in its study population.(6) Studies in the other regions also depict varying patterns of seroprevalence depending on the timelines of the pandemic in their respective countries.(7-9) One of the largest seroprevalence studies around the globe in England has reported a decline in seroprevalence estimates of IgG by 26.5% between June and September 2020, from 6.0% to 4.4% by lateral flow immunoassay (LFIA) tests.(10)

This study in Pakistan is part of the global 'WHO Unity Studies' initiative with the main objective of estimating the seroprevalence of IgG and IgM antibodies to the 'SARS-CoV-2' coronavirus in the general population in the country. This study will not only provide data regarding the exposure of the general population to COVID-19, but would also shed light on some risk factors for developing SARS-CoV-2 infection. Overall, we expect that the estimates would provide us with ample evidence to gauge the population-level scenario of COVID-19 in the country as well as provide insights into other epidemiological aspects of the disease, including the risk factors for developing SARS-CoV-2 infection.

METHODS

Patient and Public involvement Statement: Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

This population-based, age-stratified cross-sectional study, was conducted at the level of households (HH) between October 21st and November 8th, 2020, with Ethics Approval from the Institutional Ethical Review Committee of Health Services Academy, Islamabad.

This was a nationwide study in the four provinces (Punjab, Sindh, Baluchistan and Khyber Pakhtunkhwa/KP) and two regions (Azad Jammu & Kashmir (AJK) and Gilgit Baltistan (GB)) of Pakistan. Individuals of all ages and genders were eligible to participate.

Sample size calculation was done using Open Source Epidemiologic Statistics for Public Health [a] using multistage sampling design (with expected prevalence of 11% which were preliminary estimates from a previous national level study;(11)) difference between upper and lower limit of the interval estimate as 1.25% (0.75% on either side); and a design effect of '2'). The resultant sample size of 4,803 was rounded off to 5,000 and distributed equally among the ten districts under study as well as among all age brackets.

The study was conducted in ten districts of the country according to the following criteria: one high and one low prevalence district each was selected from provinces; and the highest-prevalence district each was selected from regions AJK and GB. In Pakistan's context, high prevalence was taken as a cumulative of more than 500 COVID-19 cases (Tier 1); and low prevalence as a cumulative of less than 500 cases (Tier 2) by the cut-off date of June 30th, 2020. The population of district and other characteristics were not used during sampling due to variable testing rates by population in various districts.

After district selection, Union Council (UC) selection was done with the aim of recruiting 500 households/participants randomly from each district (25 participants defined as one cluster). UCs were randomly selected from each district; the selection was in accordance with the "Urban UC: Rural UC" ratio for each district. Where there were less than 20 UCs in a district, the number of clusters were increased to make the total equal to 500 participants per district. Systematic random sampling was employed for the next stage i.e. household selection. Thereafter, one individual was randomly selected from each consenting household in line with the age distribution of the study. Each cluster of 25 had five participants from each age group. In this way 25 HH/individuals were recruited for each cluster, with random selection at different stages (as described) aiming to reduce potential sources of selection bias.

Data collection for the survey included on-the-spot recruitment of households in the selected localities. After obtaining informed consent, each participant (one individual randomly selected out of all the eligible from each household) was asked to provide information to the enumerator to fill a pre-tested questionnaire. The questionnaire was adopted from WHO Unity Studies protocol and collected information of socio-demographic variables, medical and symptom history, preventive behaviours (note: handwashing was defined as follows- individuals were inquired about number of times they had washed their hands with soap for 20 seconds; the variable was categorized into those washing hands at least six times and those less than six times.), complications and history of recent death in family followed by Rapid Diagnostic Test (RDT) results performed by a trained phlebotomist. (3) RDT was performed with 'Bioperfectus' kits for IgG/IgM, and the results were provided to each participant on the spot (within a few minutes). Prior to the survey, data collectors and field teams underwent training, which included PPE usage

[a] https://www.openepi.com/Menu/OE_Menu.htm

for infection prevention and control. Monitoring visits were also conducted during the data collection period to enhance data quality and results. We computed descriptive statistics and univariate logistic regression that examined associations of seropositivity with age, gender, location, tier, gender and symptoms with 95% confidence intervals (CI).

The multivariate binary logistic regression model

$$\log\left(\frac{p}{1-P}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \qquad (1)$$

where p is the probability that an individual is seropositive, β_0 is the intercept, β_i are the coefficients and X_i represents the independent variables. The outcome variable represents the seropositivity which is a binary categorical variable whereas location, age, mask use and contact with COVID-19 positive person are the predictor variables. A forward-stepwise process was utilized for the selection of significant variables in the final model. During the analysis, some interaction terms were considered but not included in the final model because they were not statistically significant. The selection of multivariate logistic model was based on Homer-Lemeshow goodness of fit, biological interpretability and statistical significance. The significance level or alpha was 0.05. Data analysis for the survey was performed using SPSS version 23.

RESULTS

During field work, recruitment was continued till the target sample was achieved. A total of 6,599 households were reached for on-the-spot participation, of which 1,601 (24%) did not provide consent to participate. A total of 4,998 households (with one individual per household) consented to participate in the study across the ten selected districts of Pakistan. Individuals of all ages and genders were eligible. The proportion of males in the recruited participants was 51%. The mean age of males (31.7 years) was similar to that of females (32.8 years). Almost two thirds (62%) of individuals were recruited from rural areas. Male to female ratio of sampled individuals was similar for rural and urban areas.

Information on seroprevalence was available for all 4998 participants with no missing values. The overall seroprevalence of SARS-CoV-2 antibodies was 7.1%. Almost 6.3% of individuals were IgG positive while IgM positivity was 1.9%. Seroprevalence in districts ranged from 0.0% (Ghotki) to 17.0% (Gilgit). Most of the districts reported a range of 6% to 9% (Table 1). The seroprevalence among different age groups ranged from 4% (0-9 years) to 10% (40-59 years) (Table 2). A total of 4% reported to have had contact with a COVID-19 positive individual.

Table 1 : Seroprevalence at District Level

District	Reported prevalence based on RT-PCR[a]	Positive (%)	Negative (%)
Quetta	High	16(3.2)	484(96.8)
Mardan		31(6.2)	470(93.8)
Rawalpindi		34(6.8)	467(93.2)
Ghotki		0 (0.0)	498(100)
Muzaffarabad	Low	39(8.6)	460(92.2)
Gilgit		85(17.0)	415(83.0)
Sibbi		24(4.8)	474(95.2)
Abbottabad		42(8.4)	460(91.6)
Lodhran		43(8.6)	457(91.4)
Jacobabad		37(7.4)	462(92.6)

Table 2 Seroprevalence by age, place, district tier, gender and (significant)symptoms

Variables		Positive (%)	Negative (%)	Odds ratio (95% CI)	p-value
Seroprevalence for COVID-19		351(7.1)	4647(93.0)		
Gender (*n*=4,998)	Male	178(6.9)	2394(93.1)	0.97(0.78-1.20)	0.879
	Female	173(7.1)	2253(92.9)		
Age (years) (*n*=4,997)	0 – 9	38(3.9)	935(96.1)		
	10 – 19	46(4.5)	973(95.5)	0.43(0.29-0.63)	0.000
	20 – 39	80(7.8)	950(92.2)	0.49(0.34-0.71)	0.000
	40 – 59	101(10.1)	899(89.9)	0.88(0.64-1.21)	0.390
	60+	85(8.7)	890(91.3)	1.18(0.87-1.59)	0.309
Location (*n*=4,998)	Urban	154(8.1)	1741(91.9)	1.30(1.04-1.62)	0.014
	Rural	197(6.3)	2906(93.7)		
District prevalence (reported cases) (*n*=4,998)	Low prevalence (tier-2)	270(9.0)	1919(96.0)	2.34(1.81-3.02)	0.000
	High prevalence (tier-1)	81(4.1)	2728(91.0)		
Contact with COVID-19 positive case (*n*=4,971)	Yes	25(13.9)	155(86.1)	2.34(1.81-3.03)	0.001
	No	292(6.7)	4041(93.3)		
Sore throat (*n*=4,987)	Yes	118(10.6)	998(89.4)	1.86(1.46-2.32)	0.000
	No	233(6.0)	3638(94)		
Fatigue (*n*=4,991)	Yes	43(10.9)	351(89.1)	1.706(1.22-2.39)	0.002
	No	308(6.7)	4289(93.3)		

Joint ache (*n*=4,988)	Yes	65(10.7)	541(89.3)	1.734(1.31-2.30)	0.000
	No	284(6.5)	4.98(93.5)		
High grade fever (*n*=4,986)	Yes	123(8.6)	1313(91.4)	1.365(1.09-1.72)	0.008
	No	228(6.4)	3322(93.6)		
Cough (*n*=4,993)	Yes	98(8.5)	1055(91.5)	1.317(1.03-1.67)	0.026
	No	253(6.6)	3587(93.4)		
Runny nose (*n*=4,987)	Yes	97(8.5)	1044(91.5)	1.32(1.03-1.7)	0.026
	No	253(6.6)	3593(93.4)		

The use of preventative behaviors was also studied. The use of face-masks while going out in public was reported to be 63%. Mask use was similar in urban and rural areas (63%). It was highest in 20-59 years age group (68%), while 60+ group reported relatively less use (49%). Mask use increased incrementally with education, from 36% in those non-educated to 82% in individuals above matric (ten years of education). Handwashing (washing hands at least six times with soap and water for 20 seconds in last 24 hours) was reported relatively less compared to mask use (39%). Handwashing was higher in urban (44%) areas; among females (43%); and increased with education, being highest in individuals above matric (Table 3).

Table 3 Mask use and handwashing practices by socio-demographic characteristics

Variables		**Mask use (%)**			**Handwashing (%)[a]**		
		Yes	**No**	**Total(%)**	**Yes**	**No**	**Total(%)**
Overall		3,128(62.6)	1,844(37.1)		1,946(38.9)	3,052(61.1)	
Location	Urban	1,185(62.9)	699(37.1)	1,884(37.8)	839(44.3)	1,056(55.7)	1,895(37.9)
	Rural	1,943(62.9)	1,145(37.1)	3,088(62.1)	1,107(35.7)	1,996(64.3)	3,103(62.1)
District Tiers	Tier-1	1,121(56.5)	864(43.5)	1,985(39.9)	762(38.1)	1,238(61.9)	2,000(40.0)
	Tier-2	2,007(67.2)	980(32.8)	2,987(60.0)	1,184(39.5)	1,814(60.5)	2,992(60)
Gender	Male	1,654(64.6)	907(35.4)	2,561(51.5)	905(35.2)	1,667(33.4)	2,572(51.5)
	Female	1,474(61.1)	937(38.9)	2,411(48.5)	1,041(42.9)	1,385(27.7)	2,426(48.5)
Age	0-19	576(59.5)	392(40.5)	968(24.5)	264(27.1)	709(72.9)	973(24.5)
	20-59	1,364(67.5)	657(32.5)	2,021(51.5)	934(46.0)	1,096(54.0)	2,030(51.0)
	60+	476(49.1)	50.9(50.9)	970(24.5)	382(39.2)	593(2398)	975(24.5)
Education	No education	412(36.0)	731(39.0)	1,143(36.9)	326(28.4)	822(71.6)	1,148(36.9)
	Primary	194(61.0)	124(24.8)	318(10.3)	137(42.9)	182(57.1)	319(10.3)
	Matric	491(75.2)	162(18.4)	653(21.2)	307(46.5)	353(53.5)	660(21.2)
	Above matric	797(81.6)	180(64.0)	977(36.9)	573(58.4)	408(13.1)	981(31.6)

[a]

Symptoms during past three months were inquired to look for possible association with COVID-19 seropositivity. Total of 23 symptoms were inquired relating to multiple systems. The symptoms shown in **Error! Reference source not found.** were significantly higher in seropositive individuals. Sore throat, fatigue, and joint aches were strongly associated with seropositivity. Among COVID-19 seropositive individuals, 68% had at least one symptom in last two months, while 32% reported to be completely asymptomatic during this period.

Almost 24% of individuals reported having at least one comorbidity. Hypertension was reported the most (18%), followed by diabetes (5%) and chronic kidney disease (2%). The reported occurrence of heart disease (1%) and asthma (2%) was relatively lower. Reported prevalence of at least one comorbidity increased with age with maximum being reported for 60+ age group (54%), followed by 40-59 years age group (40%).

Multivariate logistic regression analysis was performed to identify factors associated with seropositivity (Table 4). Urban residents were more likely to test positive for COVID-19 antibodies than rural residents (OR 1.29, 95% CI 1.04 - 1.61). Individuals aged 20 and above were about twice as likely to be seropositive than those who were 0-9 years old. Odds of seropositivity were also high among individuals who did not wear face mask (OR 1.54, 95% CI 1.20 – 1.975) and in those who reported contact with COVID-19 person (OR 1.81, 95% CI 1.16 – 2.83).

For any discrepancy and to ensure completeness of data, field teams were contacted and cross-checked. All cases are included in the analysis and missing data up to 1% were considered acceptable because of low occurrence.

Table 4 Factors associated with seropositivity

Variables		**b**	**S.E.**	**Adjusted OR (95% CI)[a]**	**p-value**
Location	Urban	0.256	0.113	1.29(1.035-1.612)	0.024
	Rural[b]				
Age (years)	0 – 9[b]				
	10 – 19	0.112	0.225	1.11(0.7-1.737)	0.618
	20 – 39	0.650	0.204	1.92(1.285-2.854)	0.004
	40 – 59	0.975	0.197	2.65(1.803-3.899)	0.000
	60+	0.894	0.202	2.45(1.646-3.630)	0.000
Mask use	Yes	0.434	0.126	1.54(1.205-1.975)	0.001
	No[b]				
Contact with COVID-19 positive case	Yes	0.596	0.227	1.81(1.163-2.831)	0.009
	No [b]				
Constant		-3.599			

[a] Hosmer-Lemshow goodness-of-fit test: chi-square value = 8.322, P = 0.403
[b] Reference category

DISCUSSION

The survey through the use of Unity Studies' age-stratified approach, estimated the national seroprevalence of IgG and IgM antibodies for COVID-19 to be 7.1% in Pakistan based on RDT testing. Among the included districts, the highest prevalence was observed for Gilgit, followed by Lodhran and Muzaffarabad (Table 1). This study was initiated in the last week of October 2020 and field activities were completed by second week of November 2020. This was a time frame, when the first wave of the pandemic was considered to have largely subsided, the new number of cases per day was markedly lower, and there was a threat of a second wave of the pandemic in the forthcoming winter months of November onwards. During this time frame < 1000 confirmed cases per day were being recorded, in the backdrop of the highest daily number of cases (6,825) that had been reported on 13th June, 2020.(2)

Although other tests such as ELISA (Enzyme-linked immunosorbent assay) offer greater accuracy than RDTs in terms of antibody detection,(12) due to practical and operational issues, RDTs were opted for this large-scale population-based study to estimate the seroprevalence of antibodies to COVID-19, with the results for on-the-spot testing in the field available within twenty minutes. Some researchers using similar kits to those used in this survey reported the sensitivity (to detect IgG/IgM)to range between 41% (at 1-5 days since symptom onset in patients positive by RT-PCR) to 100% (at > 20 days since symptom onset in patients positive by RT-PCR); while the reported specificity when compared with PCR was 95%(13) This time-dependent sensitivity of RDTs to detect SARS-Cov-2 antibodies has also been noted elsewhere.(14) Due to the widely varying sensitivity of the testing method, it is likely that the actual seroprevalence may have been much higher than this study's estimates.

It should be noted that the districts' population size and the age structure of the districts' population were not taken into account during district selection or selection of individuals of different age groups from within households, this may have potentially introduced bias in the seroprevalence estimates.

Additionally, 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 limitations.

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

About one third of seropositive individuals had reported to have experienced symptoms during the past two months. Six out of the studied 23 symptoms experienced during the past two months were found to be significantly associated with seropositivity in the univariate analysis, most of which were respiratory/pharyngeal symptoms including sore throat, shortness of breath, cough and runny nose. High grade fever, joint aches and fatigue were the three generalized symptoms associated with seropositivity. Although the symptoms were self-reported with a possibility for recall bias, information on symptoms was obtained before the testing was done hence it is likely that any misclassification may have been non-differential.

Similar to what has been reported in some other national studies (including a large-scale national-level household study by Pollán et al in Spain), gender was not found to be significantly associated with seropositivity for COVID-19 antibodies.(8, 9, 15) The association of age with seropositivity increased with age until 59 years, and declined slightly in those above this age bracket. As observed previously, nasal gene expression of angiotensin-converting enzyme 2 (ACE2) has been postulated to be responsible for this age-related pattern, and children have been reported to be less susceptible to contracting COVID-19 than adults potentially due to the role of innate immunity(16); and the protective effect of lower levels of this enzyme in their nasal epithelium, as also reported elsewhere.(17, 18)

The risk of seropositivity doubled in those with a history of exposure to a diagnosed COVID-19 patient and was found to be statistically significant. A study in Italy reported an even higher Odds ratio of 2.5 in those who had previous contact with a case.(19) Association of COVID-19 with household contact with known cases of COVID-19 has also found to be significant in other prevalence studies.(20)

Differences in seroprevalence between Tier 1 (districts considered to be high transmission areas by the end of June 2020) and Tier-2 districts (districts considered to be low transmission areas at the end of June 2020), were significant at about 4.9%. However, it is important to note that the seroprevalence was lower in Tier 1 districts in this study. This pattern suggests that the transmission scenarios would have evolved in three-four months since the chosen cut-off date, and the areas earlier considered to be higher transmission at the end of June (i.e. Tier 1 districts) may have now become areas of relatively lower transmission, and vice versa. The said difference is understood to be possibly due to a higher proportion of population in previously high risk districts, having possibly experienced the exposures, with possibility of reduction in IgM levels across a time span of 90 or more days (i.e. the time span between June and October 2020).

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 2.9 million 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 kilometer)

Urban and rural areas were sampled from each district in accordance with the urban to rural ratio for that particular district. Overall, urban areas reported a higher prevalence (8.1%) than rural areas (6.3%), and urban residents were more likely to be seropositive for COVID-19 than rural residents (Table 4). While some researchers have argued that larger city sizes tend to have higher attack rates, (21) other studies have gone further so as to report that while urban areas do have a propensity for earlier outbreaks than rural areas, population density is not significantly associated with COVID-19 cases.(22) Our results seem to align with the former, i.e. showing a significant association of urban residence with COVID-19 seropositivity.

Among studied behaviors, use of masks was reported by about two-thirds of the study participants with similar values in rural and urban areas. Mask use was found to be significantly linked with seropositivity in logistic regression in our study (Table 4), in line with wide-ranging evidence.(1, 23, 24) On the other hand, hand-washing(at least six times per day) had a lower prevalence (39%) in the study population and was not found to be a significant risk factor in this study, although other researchers have reported protective benefits of hand hygiene.(25) Since these have both been recommended preventive behaviours during the pandemic, the results may have been affected by Social Desirability Bias(26) in the study population, causing over-reporting of these behaviours. In such a case, any true association of these behaviours with seropositivity for COVID-19 may have been masked, if they do indeed influence seropositivity for SARS-CoV-2.

Various studies to estimate the seroprevalence of COVID-19 have been conducted in the country at different scales. For example, preliminary results from a national sero-prevalence study that was conducted during July 2020 reflected a prevalence of 11.2%.(11) Seroprevalence studies at a smaller scale in the country have also been conducted in Karachi city and Islamabad Capital Territory(ICT), with the former reporting figures of 9.7% and 15.1% in 'low-transmission' and 'high-transmission' areas in the city, respectively,(27) and the latter reporting an overall seroprevalence of 14.5%.(28, 29) Results from the aforementioned sub-national surveys differ from those of the current study in which the seroprevalence was found to 4.1% in Tier-1 districts and 9.2% in Tier-2 districts, respectively. This may have been due the differing time frames of the studies in Karachi and ICT, which were conducted during the first wave of COVID-19 in the country; while the current study was conducted during October-November 2020, when the first wave of the pandemic had largely subsided.

Overall, it is likely that the seroprevalence estimates may have been affected by low sensitivity of the testing methods(causing underestimation of the values); and the estimates for association of seropositivity with risk factors for developing SARS-CoV-2 infection may have been affected by biases including recall and social desirability bias among the study population. The seroprevalence estimates provided by this study may be interpreted with caution as an estimated value for the general population, particularly since the age structure of the study population differs from the demographic distribution of the population in the country.

CONCLUSIONS

Notwithstanding the limitations of the study, this survey provides useful prevalence estimates as well as information on risk factors for developing SARS-CoV-2 infection. The results also show that the youngest age groups have the lowest proportion of seropositivity as compared to those aged 40 years and above. Interestingly, Tier 1 districts (considered to be high risk based on the number of PCR test based confirmed cases by the end of June 2020) reflected lower prevalence as compared to Tier 2 districts which may perhaps be depicting reversing patterns at the population level. A history of contact with a confirmed COVID-19 case, being an urban resident and mask use were key risk factors for developing SARS-CoV-2 infection. Keeping view of these findings, it is premised that similar studies need to be replicated at the population level on a regular basis to monitor the disease and immunity patterns related to COVID-19.

List of Abbreviations

COVID-19- The Coronavirus disease of 2019

WHO- World Health Organization

ELISA - Enzyme-linked immunosorbent assay

RDT - Rapid Diagnostic Test

SARS-CoV-2 - Severe acute respiratory syndrome coronavirus 2

RT-PCR - Reverse transcription polymerase chain reaction

ACE2- Angiotensin-converting enzyme 2

HH- Household

OR- Odds Ratio

IRB - Institutional Review Board

KP - Khyber Pakhtunkhwa

AJK- Azad Jammu & Kashmir

GB- Gilgit Baltistan

UC- Union Council

LFIA - Lateral flow immunoassay

- **Competing interests**

  The authors declare that they have no competing interests

- **Funding**

This study was funded by Health Services Academy (HSA), Islamabad which is affiliated with the Ministry of National Health Services Regulations and Coordination, Government of Pakistan. HSA was involved in all aspects of the survey including study design/technical supervision; collection, analysis and interpretation of the data; in the writing of the manuscript; and in the decision to submit the paper for publication.

Funding/grant award number: F-11-2020-WHO/DAi/HSA

- **Authors' contributions**

AH and AA supervised the overall process; AA contributed to the analysis, manuscript writing and monitoring of data collection; KS led the coordination of all survey activities including design, training, administrative aspects and data analysis; led the process of preparing tools for data collection; contributed to training of data collectors and monitoring of the data collection process, and to the analysis and reporting of Results; MM supported preparation of data collection tools, contributed to training of data collectors, drafted the Household sampling methodology, conducted household refusal data analysis, contributed to monitoring of data collection, conducted literature reviews to analyse epidemiological data from the survey, and prepared the initial manuscript draft; MU contributed to data management, analysis and reporting. FA supported preparation of data collection tools, contributed to monitoring of data collection and training of data collectors. All authors reviewed and approved the manuscript.

- **Acknowledgements**

Not Applicable

- **Data Sharing Statement**
Data are available upon reasonable request

These are de-identified participant data, available from Health Services Academy(HSA) (Dr. Shahzad Ali Khan, Vice Chancellor, HSA. Email: shahzad@hsa.edu.pk). Data may be reused for future meta-analysis upon reasonable justification of benefit at regional/global level after permission. Additional information is also available (protocol inclusive of study tool, and statistical analysis plan).

- **Ethics**
This study involves human participants and was approved by an Ethics Committee(s) or Institutional Board(s)

Please state the name of the Ethics Committee(s) or Institutional Board(s) that approved this study with reference number or ID for your ethics approval OR the reason one was not provided:

Institutional Ethical Review Committee of Health Services Academy, Ministry of National Health Services, Islamabad, Pakistan

No. 7-82/IERC-HSA/2020-33

References

1. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020 2020 [cited 2020 22 December]. Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
2. Government of Pakistan. COVID-19 Dashboard 2020 [cited 2020 22 December]. Available from: http://covid.gov.pk/stats/pakistan?locale=en.
3. World Health Organization. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigation Protocols 2020 [cited 2020 22 December]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations.
4. Arora RK, Joseph A, Van Wyk J, Rocco S, Atmaja A, May E, et al. SeroTracker: a global SARS-CoV-2 seroprevalence dashboard. The Lancet Infectious Diseases. 2020.
5. Poustchi H, Darvishian M, Mohammadi Z, Shayanrad A, Delavari A, Bahadorimonfared A, et al. SARS-CoV-2 antibody seroprevalence in the general population and high-risk occupational groups across 18 cities in Iran: a population-based cross-sectional study. The Lancet Infectious Diseases. 2020.
6. Murhekar MV, Bhatnagar T, Selvaraju S, Rade K, Saravanakumar V, Thangaraj JWV, et al. Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, May-June 2020. Indian Journal of Medical Research. 2020;152(1):48.
7. Xu X, Sun J, Nie S, Li H, Kong Y, Liang M, et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. Nature Medicine. 2020:1-3.
8. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. The Lancet. 2020;396(10250):535-44.
9. Herzog S, De Bie J, Abrams S, Wouters I, Ekinci E, Patteet L, et al. Seroprevalence of IgG antibodies against SARS coronavirus 2 in Belgium: a prospective cross-sectional study of residual samples. medRxiv. 2020.
10. Ward H, Cooke G, Atchison CJ, Whitaker M, Elliott J, Moshe M, et al. Declining prevalence of antibody positivity to SARS-CoV-2: a community study of 365,000 adults. MedRxiv. 2020.
11. Dawn News. 11pc Pakistanis have developed protective immunity: study 2020 [cited 2020 22 December]. Available from: https://www.dawn.com/news/1575607/11pc-pakistanis-have-developed-protective-immunity-study.
12. Kontou PI, Braliou GG, Dimou NL, Nikolopoulos G, Bagos PG. Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis. Diagnostics. 2020;10(5):319.
13. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, et al. Evaluation of SARS-CoV-2 serology assays reveals a range of test performance. Nature Biotechnology. 2020;38(10):1174-83.
14. Rashid ZZ, Othman SN, Samat MNA, Ali UK, Wong KK. Diagnostic performance of COVID-19 serology assays. The Malaysian Journal of Pathology. 2020;42(1):13-21.
15. Lai C-C, Wang J-H, Hsueh P-R. Population-based seroprevalence surveys of anti-SARS-CoV-2 antibody: An up-to-date review. International Journal of Infectious Diseases. 2020.
16. Yang HS, Costa V, Racine-Brzostek SE, Acker KP, Yee J, Chen Z, et al. Association of age with SARS-CoV-2 antibody response. JAMA network open. 2021;4(3):e214302-e.
17. Patel AB, Verma A. Nasal ACE2 levels and COVID-19 in children. Jama. 2020;323(23):2386-7.
18. Bunyavanich S, Do A, Vicencio A. Nasal gene expression of angiotensin-converting enzyme 2 in children and adults. Jama. 2020.

19. Vena A, Berruti M, Adessi A, Blumetti P, Brignole M, Colognato R, et al. Prevalence of antibodies to SARS-CoV-2 in Italian adults and associated risk factors. Journal of clinical medicine. 2020;9(9):2780.
20. Menachemi N, Yiannoutsos CT, Dixon BE, Duszynski TJ, Fadel WF, Wools-Kaloustian KK, et al. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample—Indiana, April 25–29, 2020. Morbidity and Mortality Weekly Report. 2020;69(29):960.
21. Stier A, Berman M, Bettencourt L. COVID-19 attack rate increases with city size. Mansueto Institute for Urban Innovation Research Paper Forthcoming. 2020.
22. Carozzi F. Urban density and COVID-19. 2020.
23. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. The Lancet. 2020.
24. Gandhi M, Beyrer C, Goosby E. Masks do more than protect others during COVID-19: reducing the inoculum of SARS-CoV-2 to protect the wearer. Journal of general internal medicine. 2020;35(10):3063-6.
25. Ma QX, Shan H, Zhang L, Li M, Yang M, Chen JM. Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2.
26. Grimm P. Social desirability bias. Wiley international encyclopedia of marketing. 2010.
27. Nisar MI, Ansari N, Amin M, Khalid F, Hotwani A, Rehman N, et al. Serial population based serosurvey of antibodies to SARS-CoV-2 in a low and high transmission area of Karachi, Pakistan. medRxiv. 2020.
28. Dawn News. 300,000 persons infected with Covid-19 in Islamabad: survey 2020 [cited 2020 22 December]. Available from: 300,000 persons infected with Covid-19 in Islamabad: survey
29. Pakistan Go. COVID-19 Tracker (Press Release, July 2020) 2020 [cited 2020 22 December]. Available from: http://covid.gov.pk/press-release-details/2.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	3
Objectives	3	State specific objectives, including any prespecified hypotheses	3
Methods			
Study design	4	Present key elements of study design early in the paper	4, 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4, 5
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	4, 5
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	4-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	4, 5
Bias	9	Describe any efforts to address potential sources of bias	4, 5
Study size	10	Explain how the study size was arrived at	3, 4
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5- 8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7, 8
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	8
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of 0cases and controls was addressed	NA

Cross-sectional study—If applicable, describe analytical methods taking account of sampling strategy	
(*e*) Describe any sensitivity analyses	NA

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (e.g. demographic, clinical, social) and information on exposures and potential confounders	5, 6
		(b) Indicate number of participants with missing data for each variable of interest	5, 6
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	
		Cross-sectional study—Report numbers of outcome events or summary measures	5-8
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	6-8
		(*b*) Report category boundaries when continuous variables were categorized	6-7 (Tables 2 & 3)
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
Other analyses	17	Report other analyses done—e.g. analyses of subgroups and interactions, and sensitivity analyses	
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	9-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results	11
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.