

Journal Pre-proof

Serial population-based sero-surveys for COVID-19 in two neighborhoods of Karachi, Pakistan

Muhammad Imran Nisar, Nadia Ansari, Farah Khalid, Mashal Amin, Hamna Shahbaz, Aneeta Hotwani, Najeeb Rehman, Sierra Pugh, Usma Mehmood, Arjumand Rizvi, Arslan Memon, Zahoor Ahmed, Ashfaqe Ahmed, Junaid Iqbal, Ali Faisal Saleem, Uzma Bashir Aamir, Daniel B Larremore, Bailey Fosdick, Fyezah Jehan



PII: S1201-9712(21)00258-7

DOI: <https://doi.org/10.1016/j.ijid.2021.03.040>

Reference: IJID 5242

To appear in: *International Journal of Infectious Diseases*

Received Date: 13 January 2021

Revised Date: 7 March 2021

Accepted Date: 11 March 2021

Please cite this article as: Nisar MI, Ansari N, Khalid F, Amin M, Shahbaz H, Hotwani A, Rehman N, Pugh S, Mehmood U, Rizvi A, Memon A, Ahmed Z, Ahmed A, Iqbal J, Saleem AF, Aamir UB, Larremore DB, Fosdick B, Jehan F, Serial population-based sero-surveys for COVID-19 in two neighborhoods of Karachi, Pakistan, *International Journal of Infectious Diseases* (2021), doi: <https://doi.org/10.1016/j.ijid.2021.03.040>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Title: Serial population-based sero-surveys for COVID-19 in two neighborhoods of Karachi, Pakistan

Running Title: COVID-19 Sero-survey in Karachi, Pakistan

Authors: Muhammad Imran Nisar MBBS, MSc; ^a Nadia Ansari MBBS, MSc; ^a Farah Khalid MSc; ^a Mashal Amin BDS, MSc; ^a Hamna Shahbaz MBBS, ^a Aneeta Hotwani MSc; ^a Najeeb Rehman BSc; ^a Sierra Pugh MSc; ^b Usma Mehmood, MSc; ^a Arjumand Rizvi MPhil; ^a Arslan Memon MBBS, ^c MPH; Zahoor Ahmed MBBS; ^c Ashfaque Ahmed MBBS; ^c Junaid Iqbal PhD; ^a Ali Faisal Saleem MBBS, MSc; ^a Uzma Bashir Aamir MBBS, ^d PhD; Daniel B Larremore PhD; ^e Bailey Fosdick PhD; ^b Fyezah Jehan MBBS, MSc ^a

Affiliations:

^a Department of Paediatrics and Child Health, Aga Khan University, Stadium Road, Karachi 74800, Pakistan; Tel: +92 21 3486 4981

^b Colorado State University, Fort Collins, Colorado, USA

^c Health Department, Government of Sindh, Karachi, Pakistan

^d World Health Organization Country Office, Karachi, Sind, Pakistan

^e University of Colorado, Boulder, Colorado, USA

***Corresponding Author Bio**

Dr Fyezah Jehan is an Associate Professor and an Infectious Diseases Specialist in the Department of Pediatrics and Child Health at The Aga Khan University.

Emails: Imran.nisar@aku.edu, Nadia.ansari@aku.edu, farah.khalid@aku.edu,
mashal.amin@aku.edu, hamna.shahbaz@aku.edu, aneeta.hotwani@aku.edu,
Najeeb.rehman@aku.edu, sierra.pugh@colostate.edu; usma.mehmood@aku.edu;
arjumand.rizvi@aku.edu, arslanmemon213@gmail.com; dhoeastkarachi@gmail.com;
Junaid.iqbal@aku.edu; ali.saleem@aku.edu; bashiru@who.int;
daniel.larremore@colorado.edu ; Bailey.Fosdick@colostate.edu ; Fyezah.jehan@aku.edu;

Highlights:

- Prevalence of SARS-CoV-2 is high in Karachi
- Most sero-positive cases are reported to be asymptomatic
- Contact within households is linked with high probability of infection transmission
- Seroprevalence studies serve as efficacious tools to determine spread of infection

Abstract- 209 words

Objective: To determine population-based estimates of COVID-19 in a densely populated urban community of Karachi, Pakistan.

Methods: Three cross-sectional surveys were conducted in April, June and August in low- and high-transmission neighborhoods. Participants were randomly selected to provide blood for Elecsys® immunoassay for detection of anti-SARS-CoV-2 antibodies. Bayesian regression model was used to estimate seroprevalence after adjusting for the demographic characteristics of each district

Results: We enrolled 3005 participants from 623 households. In Phase 2, adjusted seroprevalence was estimated as 8.7% (95% CI 5.1-13.1) and 15.1% (95% CI 9.4 -21.7) in low and high transmission areas respectively, compared to 0.2% (95% CI 0-0.7) and 0.4% (95% CI 0 - 1.3) in Phase 1. In Phase 3, it was 12.8% (95% CI 8.3 – 17.7) and 21.5% (95% CI 15.6-28) in low and high transmission areas, respectively. CRI was 0.31 (95% CI 0.16-0.47) and 0.41(95% CI 0.28-0.52) in low and high transmission neighborhoods respectively in Phase 2. Similar trends were observed in Phase 3. Only 5.4% of participants who tested positive for COVID-19 were symptomatic. IFR was 1.66%, 0.37% and 0.26% in Phases 1, 2 and 3 respectively.

Conclusion: Continuing rounds of seroprevalence studies will help us better understand secular trends and extent of infection during the course of the pandemic.

Keywords: Seroprevalence; Seroepidemiology; COVID-19; Antibody; Surveillance; Pakistan

Text- 3026 words

INTRODUCTION

The COVID-19 pandemic has resulted in more than 62 million confirmed cases and over 1.4 million deaths globally, case fatality rate (CFR) around 5.4% and infection fatality rate (IFR) of 0.9% (1, 2). As the world hastened to respond to the global health crisis, the pandemic revealed numerous cracks in health-care systems (3). Pakistan was amongst the first low- and middle- income countries (LMICs) to be affected by the pandemic and has reported 398,024 cases and 8,025 deaths (CFR 2.51 %) (1, 4). The CFR has been variable; as high as 5.11% (KPK) and as low as 0.71% (Gilgit Baltistan) with 1.82%, 1.80% and 1.28% reported for Punjab, Sindh, and Baluchistan respectively (5).

The demographic characteristics of Pakistan are typical of most LMICs with 41% of the population under the age of 15 (6). The societal construct in Pakistan boasts a patrilineal joint family system, intergenerational co-residence of family members and average household sizes of 6 or greater (7). This is important because recent evidence indicates that households with individuals above 60 years of age are at risk of life-threatening manifestations of the disease (8). Karachi is home to people from varying diasporas across Pakistan, and the epicenter of the epidemic since 26th Feb 2020, reporting the largest number of cases in Pakistan (84,232) and accounting for 28% of all COVID-19 cases in the country (9). Crowded neighborhoods and urban slum dwellings along with poor adherence to mitigation measures in the city may have expeditiously increased infection transmission (10).

Surveillance systems form the basis for tracking COVID-19 cases, testing populations at risk and carrying out contact tracing, thereby, representing the key components of the public health response (11). Facility-based surveillance efforts are likely to miss mild and asymptomatic cases, as bolstered by evidence from a WHO-China Joint Commission Report

and several published studies which indicate that between 5-80% of sero-positive patients are noted to be asymptomatic. Therefore, household targeted serological testing can decrease biases arising from selective testing and generate concrete evidence on the role that asymptomatic transmission of infection plays in rapidly increasing infection rates (12). Household transmission is of special concern in congested neighborhoods of metropolitan cities when lockdown measures are in place. Secondary transmission from index cases in households using prospective follow-up and active symptom monitoring with nasopharyngeal polymerase chain reaction (NP-PCR) has indicated household attack rates as high as 32.4% (95% CI 22.4%–44.4%) (13). However, this exercise is resource intensive and transmission may differ between symptomatic and asymptomatic households, as symptomatic individuals are more likely to transmit the virus (14-17).

Using seroprevalence data, the conditional risk of infection (CRI), namely, the probability that an individual in a household is infected given that another household member is infected, can serve as a related index of infection within household (14, 18). Additionally, we can calculate true estimates of Infection Fatality Rate (IFR) as we are able to detect asymptomatic and mild cases who did not seek treatment.

In this study we estimate changes in seroprevalence in low- and high-transmission neighborhoods of Karachi between April and August 2020 through serial cross-sectional surveys by adapting the World Health Organization (WHO) Unity protocol. We present the estimates by age and gender and apply these to compute conditional risk of infection and infection fatality ratio. (19)

MATERIALS and METHODS

Study Participants and sample collection

We conducted the study in two areas. Four sub-administrative units (union councils) of District East were selected as high-transmission areas based on the number of cases per million reported by the provincial government. (Figure 1). One union council of District Malir was selected as a low transmission site. The study was approved by the Aga Khan University Ethical Review Committee (AKU ERC).

Three cross-sectional surveys were sequentially performed at the household level, between April 15-25, June 25- July 11 and August 17-22, 2020. Four research teams, each comprising of one data collector and one phlebotomist collected data and serology samples. A detailed line-list of cases was available in District East which allowed households to be selected through systematic random sampling as follows: A case was randomly identified from the line listing that served as a reference point. From here, the direction was determined by spinning a bottle or a pen and the *n*th interval between structures was identified using the second last digit of a bank note. The reference household was not included in the survey. In case of household refusal, the next household was approached. Facilities like hospitals, care homes, educational institutions and prison were excluded. However, health care workers within a household were not excluded. In District Malir, where we had a line listing of all residential households, a simple random sampling was done. In both areas, all household members were eligible to participate irrespective of their infection status. Approval from the household head and written informed consent or assent from individual participants was obtained.

All team members underwent training in the use of PPE, hand hygiene and safe transportation of biological samples. A 3-mL sample of blood from infants and 5- mL from older

participants, was collected by a trained phlebotomist and transported to the Infectious Diseases Research Lab for centrifugation, serum separation and storage at -20°C . Information was collected from all the participants on age, gender, occupation and household size, along with details of travel history and exposure to COVID-19 patients. Reported comorbidities, presence of symptoms, history of hospitalization we also recorded. An occupational history was also inquired from individual participants regarding working from home and otherwise. For symptoms, the clinical history recorded details of the presence of fever, respiratory symptoms such as sore throat, shortness of breath and chest pain in previous two months (Supplemental Appendix 3).

Laboratory analysis

A commercial Elecsys® Anti-SARS-CoV-2 immunoassay (Roche Diagnostics), targeting combined IgG and IgM against SARS-CoV-2 was performed at the Nutritional Research Laboratory (NRL) at Aga Khan University. The manufacturer reported a specificity greater than 99.8% and sensitivity of 100% for individuals with a positive PCR test at least two weeks prior, and 88.1% sensitivity for those 7-13 days after a PCR-positive test (20).

STATISTICAL ANALYSIS

Sample Size

The sample size for each phase of the survey was calculated to be 500 participants for each site. This allowed for an estimation of an age-adjusted prevalence in the range of 20-30 % at 95% confidence level, with a precision of $\pm 5\%$ and a design effect of 1.5 for household level clustering.

Data Management

Data entry was performed in duplicate on an SQL database (Structured Query Language Database) (21) and was thoroughly checked for completeness and consistency. Continuous variables such as age and household size were reported as mean and standard deviation (SD). Categorical variables like gender, occupation and symptoms were reported as frequencies and percentages. Participants reporting fever or respiratory symptoms in the last two months were categorized as symptomatic and presented as proportions. Age was also categorized.

Estimation of overall, age and gender stratified seroprevalence

Age and gender-stratified seroprevalence estimates were computed for each district and each phase independently, using a Bayesian hierarchical regression model. This approach, described in detail in the Supplementary Material, accounts for uncertainty due to finite lab validation data, (22) and produces estimates using typical choices of uninformative or weakly informative prior distributions (23, 24).

Given the correlation among household seropositivity values, model also took into account factors such as total number of members in a household and adjusted for test accuracy by modeling directly on the lab validation data reported by the test manufacturer (20).

Seroprevalence estimates by age and gender were then post stratified to adjust for the demographic makeup of the respective districts.

Estimates are expressed as posterior means and 95% equal-tailed credible intervals based on 20,000 samples from a Bayesian posterior distribution. All calculations were performed in Software R and samples from posterior distributions were obtained using Stan (25).

Household Conditional Risk of Infection (CRI) and Infection Fatality Rate analysis

Conditional Risk of Infection (CRI) is the probability that an individual is infected, conditioned on a household member being infected (26). CRI was calculated and presented as a fraction where the numerator is the total number of ordered pairs among infected individuals in the same household and whose denominator is the total number of ordered pairs in the same household in which the first individual in the pair is infected. A 95% confidence interval was estimated via bootstrap for each area by resampling households with replacement.

Age specific - Infection Fatality Rate (IFR): This measure estimates the prevalence of infection (including both asymptomatic and mildly symptomatic infections). The cumulative number of deaths due to COVID-19 infection were calculated till 14 days after sero-survey dissemination. This number was then divided by the figure obtained from multiplying the relevant population of an area with the adjusted estimate of seroprevalence (27).

Results

Participant Flow

A total of 3005 participants were enrolled across three phases from District East and District Malir (Figure 2 Panels, A, B and C). There were high refusal rates in both areas at the household level, 68%, 43% and 61 % in district East and 44%, 42% and 8% in district Malir, in phases 1, 2 and 3 respectively.. Amongst the households who consented to participate, individual participation rate was 82.3% (1000 out of 1215 eligible household members) in phase 1, 76.5% (1004 out of 1312 eligible household members) in phase 2 and 80% (1001 out of 1243 eligible household members) in phase 3. Table 1 describes the baseline demographic and clinical characteristics of the enrolled participants.

Seropositivity

In Phase 1 of the study, only 2 of 500 samples tested positive in District East, while none of the 500 participants tested positive in District Malir. In Phase 2 of the study, 100 of 500 samples (20.0%) tested positive in District East and 64 of 504 samples (12.7%) tested positive in District Malir. In phase 3 of the study, 119 of 500 samples (24 %) tested positive in District East and 79 of 501 samples (16%) tested positive in District Malir.

Adjusted Seroprevalence

In Phase 1, post-stratified seroprevalence was estimated to be 0.4% (95% CI 0%-1.3%) in District East and 0.2% (95% CI 0%-0.7%) in District Malir. In Phase 2, it was estimated to be 15.1% (95% CI 9.4%-21.7%) in District East, and 8.7% (95% CI 5.1%-13.1%) in District Malir. Lastly, in Phase 3, the rate was documented to be 21.5% (95% CI 15.6%-28%) in District East, and 12.8 % (95% CI 8.3%-17.7%) in District Malir.

Both districts showed marked and significant increase in seroprevalence between phases sequentially with a sharp rise between Phase 1 and 2 and a smaller rise between phase 2 and 3. Overall reporting of symptoms was documented to be 9.4 %. Of the total 364 participants who tested positive, 27 (7.4 %) gave a history of fever or respiratory symptoms or both in the last 2 months (Supplementary Table 1).

Conditional Risk of Infection and Infection Fatality Rates

To measure whether individuals in the same household were more likely to have similar sero-status, we computed the conditional risk of infection (CRI) for phases 2 and 3. CRI estimates were 0.41 (95% CI 0.28-0.52) and 0.38 (95% 0.27-0.52) in District East and 0.31 (95% CI 0.16-0.4) and 0.33 (95% 0.12-0.47) in District Malir in phases 2 and 3 respectively. In parallel, the age specific Infection Fatality Rate (IFR), a measurement calculated based on the cumulative total of infected persons (based on seroprevalence) and number of deaths

(obtained from local government surveillance data), was calculated to be 1.66% in phase 1, 0.37% in phase 2 and 0.26% in phase 3.

The rise in seroprevalence in district East corresponded with the epidemiology as ascertained through daily case reporting in the district as well as the four study sampling sites. (Figure 3). Seropositivity rates were indistinguishable between male and female within each district as well as between age groups (Table 2, graphically represented in supplementary Figure 1-A and 1-B).

Discussion

This study summarizes through three serial cross-sectional surveys, the seroprevalence rates in two neighborhoods of Karachi, starting from the very early phase to post peak phase of the pandemic. We report an increase in sero-prevalence over time, with only a tiny fraction of seropositive individuals reporting any symptom. A study from Karachi, Pakistan reported a seroprevalence of 17.5% among workers across various occupational sectors while another from Lahore found a prevalence of 15.6 among police force personnel (28, 29). The National Institute of Blood Disease (NIBD) in Karachi found an overall estimate of 36% among their participants that included industrial workers (50% positivity), health care workers (13% positivity), general public (15.6% positivity) and healthy blood donors (36% positivity) (30).. Preliminary reports of seroprevalence studies from international literature have highlighted high rates of seroprevalence in large metropolitan areas. A study from Mumbai, India reported up to 55% (31), while a survey carried out in the Guilan province of Iran showed a prevalence of 33% at the peak of the pandemic (32, 33).

The three serial serosurveys conducted in the same populations in Karachi indicated an initial early increase in the second phase that continued over the next few months as demonstrated by Figure 3. A number of reasons can explain this trend; the first phase of the

study was done early in the pandemic, 3 weeks after a provincial lockdown was declared while the second phase was conducted after lockdown measures were eased in anticipation of the religious festivities observed during Eid in Pakistan. Although individual associations cannot be discerned, effectiveness of social mitigation measures could explain the smaller rise in phase three.

The survey did not identify any difference in seroprevalence between males and females or any age-specific trends in infection rates across different age categories. Prevalence appeared to increase with age and was documented to be consistently high between 19-39 years of age and 40-59 years of age in both males and females. This is consistent with age related seropositivity patterns in the literature (34). Our survey found a large number of asymptomatic sero-positive individuals. Only 3 out of 10 reported any respiratory symptoms, with or without fever. In contrast to our findings, the proportion of asymptomatic infections was reported to be much lower, 27.7% (95% CI, 16.4% - 42.7%) in a meta-analysis (35). However, higher rates of asymptomatic infection have been reported in some parts of the world like India. According to WHO and the Indian Council of Medical Research (ICMR), India, the number of asymptomatic cases appears to be about 80% (36). Lower reporting of symptoms in Pakistan may be attributable to innate fears of disclosure of disease positivity and lack of awareness about the high transmissibility of the disease in the general public in the early stages of the pandemic.

The rise over time in seroprevalence even in an area of presumably low transmission indicates that seroprevalence studies may serve as important tools to determine spread of infection in populations where a large majority of the people are asymptomatic. Monitoring the general populations through serial sero-surveys can detect resurgence, especially when lockdown measures are eased and enable policy makers to devise strategies for containment of the disease. Heterogeneity between low- and high-income neighborhoods is likely and has

been suggested previously (37-39). In our study an alternate explanation for this might be different sampling strategy for the two districts. However, it is difficult to discern in which direction this will likely bias the results. Temporal trends within the districts can nevertheless be established from our study.

Our results also confirm that close contact within households is linked to a high probability of being infected and should be an important consideration in COVID-19 transmission (40). Our study indicated that the probability of an individual to acquire an infection in the presence of another infected household member, as measured by conditional risk of infection (CRI), was high between 35% - 40%. An unpublished review used a secondary attack rate, which is a more reliable indicator of intra-household transmission and reported it to be lower, i.e., 18.8% (95% CI 15.4%-22.2%) (41). CRI can hence, function as a substitute in situations where comprehensive surveillance efforts and disease notification strategy are absent and where secondary attack rates are difficult to estimate.

Our study indicated that the age specific IFR was higher in phase 1 (1.66%) and then decreased in phases 2 and 3 to 0.37% and 0.26% respectively. This may be attributed to early detection and improved management of symptomatic disease. A systematic review and meta-analysis conceded that overall IFR could be 0.68% (0.53%-0.82%) but with heterogeneity from either the studies contributing to the review or due to various factors like age and presence of co-morbidities in the population (42). Population demographics can affect estimates of population-weighted IFR, with the lowest IFRs recorded for countries with younger populations, for instance Kenya (0.09%; 95% credible interval, 0.08–0.10%) and Pakistan ((0.16%; 95% credible interval, 0.14–0.19%) (43).

The strength of our study is that it captured the baseline seropositivity at a population-level right at the beginning of the pandemic when transmission was low. Subsequent survey at two monthly intervals allowed for a sequential evaluation of changes in sero-prevalence. About one third or more of our sample also included children less than 18 years of age, focusing on an understudied age group in the pandemic. Adaptation of the standardized WHO UNITY protocol allows for comparison of results from this study with other studies that have used the same protocol.

Our study had several limitations. The geographical area of the study was limited to only two neighborhoods of Karachi, Pakistan. Although Karachi is a large metropolitan city of 16 million people which is multiethnic and socioeconomically diverse, it is definitely not representative of the whole country. Sample size did not allow for comparison between two neighborhoods. . We also had high rates of household level refusal which can potentially result in both underestimation or overestimation of prevalence rates. . Due to a limited supply chain of testing kits in Pakistan, we did not do an in-house validation on local samples, however this was compensated by modeling directly on the data reported by the manufacturer.

Conclusion:

We see increase in seroprevalence estimates in both neighborhoods of Karachi over time. Most sero-positive cases were reported to be asymptomatic. The probability of an individual to be infected following exposure within the same household was high. The infection fatality

rate decreased over time. . . Continual cross-sectional surveys are valuable in monitoring population level immunity against COVID-19 during pandemic and post vaccine era.

Journal Pre-proof

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding: This study was supported by the Infectious Disease Research Laboratory (IDRL) at the Aga Khan University in Karachi, Pakistan.

Financial Disclosure: None of the authors have any financial interest to disclose.

Acknowledgments

The authors would like to acknowledge all the data collectors, phlebotomists and laboratory personnel who made this happen in the most difficult of circumstances.

References

1. (JHU) JHU. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) 2020 [Available from: <https://coronavirus.jhu.edu/map.html>].
2. Rekatsina M, Paladini A, Moka E, Yeam CT, Urits I, Viswanath O, et al. Healthcare at the time of COVID-19: A review of the current situation with emphasis on anesthesia providers. *Best Practice & Research Clinical Anaesthesiology*. 2020.
3. Armocida B, Formenti B, Ussai S, Palestra F, Missoni E. The Italian health system and the COVID-19 challenge. *The Lancet Public Health*. 2020;5(5):e253.
4. Pakistan Go. COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination 2020 [Available from: <https://covid.gov.pk/stats/pakistan>].
5. Anser MK, Yousaf Z, Khan MA, Nassani AA, Abro MMQ, Vo XH, et al. Social and administrative issues related to the COVID-19 pandemic in Pakistan: better late than never. *Environmental Science and Pollution Research*. 2020;27(27):34567-73.
6. Mahsud-Dornan S. Pakistan, population programmes and progress. *The Ulster medical journal*. 2007;76(3):122.
7. shrinks CFs. <https://tribune.com.pk/story/1491353/census-2017-family-size-shrinks> [Census 2017: Family size shrinks]. August 26, 2017 [updated August 26, 2017. Available from: <https://tribune.com.pk/story/1491353/census-2017-family-size-shrinks>].
8. Aparicio A GS. Intergenerational Residence Patterns and COVID-19 Fatalities in the EU and the US. Bonn. Germany: IZA – Institute of Labor Economics; 2020 July 2020. Contract No.: 13452.
9. Health Department GoS. "Daily Status Report / Twitter".
10. Wasdani KP, Prasad A. The impossibility of social distancing among the urban poor: the case of an Indian slum in the times of COVID-19. *Local Environment*. 2020;25(5):414-8.
11. Organization WH. Global surveillance for COVID-19 caused by human infection with COVID-19 virus: interim guidance, 20 March 2020. World Health Organization; 2020.
12. Kumar MS, Bhatnagar T, Manickam P, Kumar VS, Rade K, Shah N, et al. National sero-surveillance to monitor the trend of SARS-CoV-2 infection transmission in India: Protocol for community-based surveillance. *The Indian Journal of Medical Research*. 2020;151(5):419.
13. Wu J, Huang Y, Tu C, Bi C, Chen Z, Luo L, et al. Household Transmission of SARS-CoV-2, Zhuhai, China, 2020. *Clinical Infectious Diseases*. 2020.
14. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature*. 2020:1-.
15. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic population. *New England Journal of Medicine*. 2020.
16. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. 2020;368(6491).
17. Streeck H, Schulte B, Kuehmerer B, Richter E, Höller T, Fuhrmann C, et al. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medrxiv*. 2020.
18. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature*. 2020;584(7821):425-9.
19. Organization WH. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigations Protocols: World Health Organization; 2020 [cited 2020. Available from: Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigations Protocols.

20. Ltd. RDI. ElectroChemiLuminescence (ECL) 2020 [Available from: <https://diagnostics.roche.com/global/en/article-listing/electroChemiLuminescence-unique-immunoassay-technology.html>].
21. The world's most popular open source database [Available from: <https://www.mysql.com/>].
22. Larremore DB, Fosdick BK, Zhang S, Grad YH. Jointly modeling prevalence, sensitivity and specificity for optimal sample allocation. *bioRxiv*. 2020.
23. Larremore DB, Fosdick BK, Bubar KM, Zhang S, Kissler SM, Metcalf CJE, et al. Estimating SARS-CoV-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. *medRxiv*. 2020.
24. Gelman A, Carpenter B. Bayesian analysis of tests with unknown specificity and sensitivity. *medRxiv*. 2020.
25. Stan - Stan 2020.
26. Curmei M, Ilyas A, Evans O, Steinhardt J. Estimating Household Transmission of SARS-CoV-2. *medRxiv*. 2020.
27. Ioannidis J. The infection fatality rate of COVID-19 inferred from seroprevalence data. *MedRxiv*. 2020.
28. Javed W, Baqar JB, Abidi SHB, Farooq W. Sero-prevalence findings from metropolises in Pakistan: implications for assessing COVID-19 prevalence and case-fatality within a dense, urban working population. *medRxiv*. 2020.
29. Chughtai OR, Batool H, Khan MD, Chughtai AS. Frequency of COVID-19 IgG Antibodies among Special Police Squad Lahore, Pakistan. *J Coll Physicians Surg Pak*. 2020;30(7):735-9.
30. Zaidi S, Rizwan F, Riaz Q, Siddiqui A, Khawaja S, Imam M, et al. Seroprevalence of anti-SARS-CoV-2 antibodies in residents of Karachi-challenges in acquiring herd immunity for COVID 19. *J Public Health (Oxf)*. 2020:fdaa170.
31. Delhi's sero-prevalence study finds 23.48 per cent people affected by COVID-19. *The Times of India*. 2020;Sect. http://timesofindia.indiatimes.com/articleshow/77085162.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst.
32. Shakiba M, Nazari SSH, Mehrabian F, Rezvani SM, Ghasempour Z, Heidarzadeh A. Seroprevalence of COVID-19 virus infection in Guilan province, Iran. *medRxiv*. 2020.
33. Bhattacharyya R, Bhaduri R, Kundu R, Salvatore M, Mukherjee B. Reconciling epidemiological models with misclassified case-counts for SARS-CoV-2 with seroprevalence surveys: A case study in Delhi, India. *medRxiv*. 2020.
34. Nickbakhsh S, Ho A, Marques DF, McMenamin J, Gunson RN, Murcia PR. Epidemiology of seasonal coronaviruses: Establishing the context for COVID-19 emergence. *The Journal of Infectious Diseases*. 2020.
35. He J, Guo Y, Mao R, Zhang J. Proportion of asymptomatic coronavirus disease 2019 (COVID-19): a systematic review and meta-analysis. *Journal of medical virology*. 2020.
36. Chatterjee S, Sarkar A, Chatterjee S, Karmakar M, Paul R. Studying the progress of COVID-19 outbreak in India using SIRD model. *medRxiv*. 2020.
37. Hooper MW, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. *Jama*. 2020.
38. de Souza WM, Buss LF, da Silva Candido D, Carrera J-P, Li S, Zarebski AE, et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. *Nature Human Behaviour*. 2020:1-9.
39. Rafael RdMR, Neto M, Depret DG, Gil AC, Fonseca MHS, Souza-Santos R. Effect of income on the cumulative incidence of COVID-19: an ecological study. *Revista Latino-Americana de Enfermagem*. 2020;28.
40. Statistics GoPBo. Housing units by number of rooms and type [Available from: www.pbs.gov.pk].

41. Madewell ZJ, Yang Y, Longini Jr IM, Halloran ME, Dean NE. Household transmission of SARS-CoV-2: a systematic review and meta-analysis of secondary attack rate. medRxiv. 2020.
42. Meyerowitz-Katz G, Merone L. A systematic review and meta-analysis of published research data on COVID-19 infection-fatality rates. International Journal of Infectious Diseases. 2020.
43. O'Driscoll M, Dos Santos GR, Wang L, Cummings DA, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature. 2020:1-6.

Figure 1. Study area and total population.

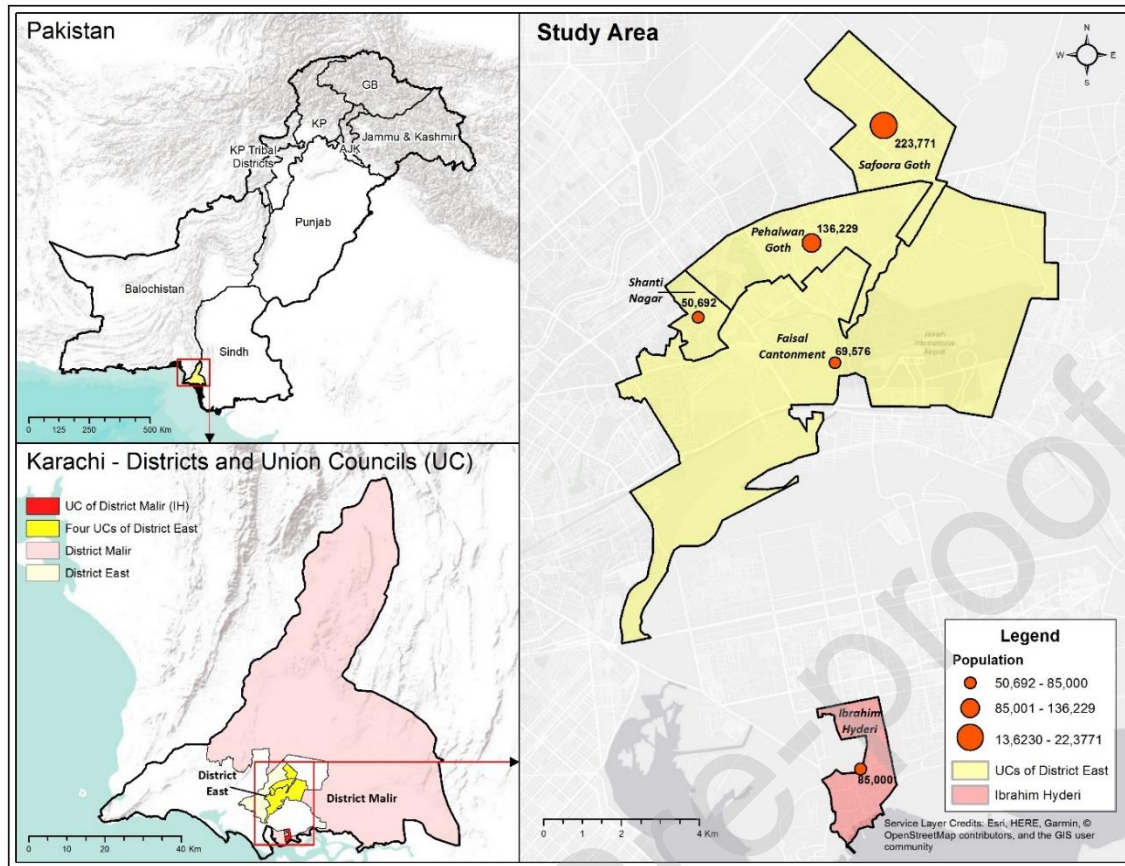


Figure 2 Panel A: Flow chart of participants in Phase 1 of study

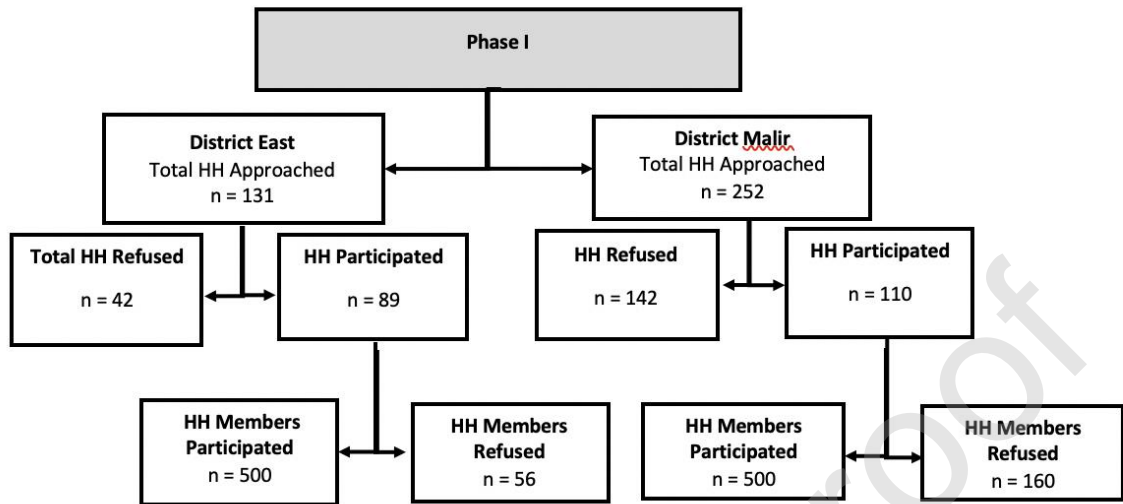


Figure 2 Panel B. Flow chart of participants in Phase 2 of study.

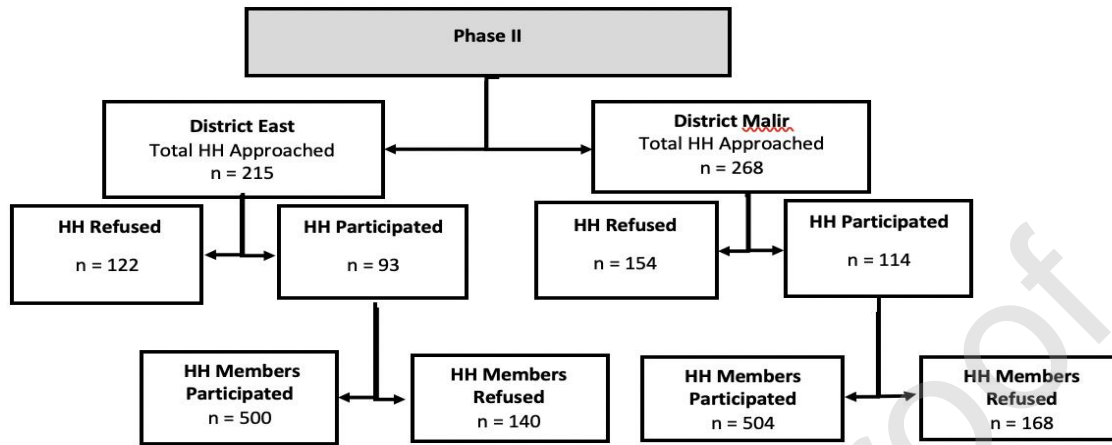


Figure 2 Panel C. Flow chart of participants in Phase 3 of study.

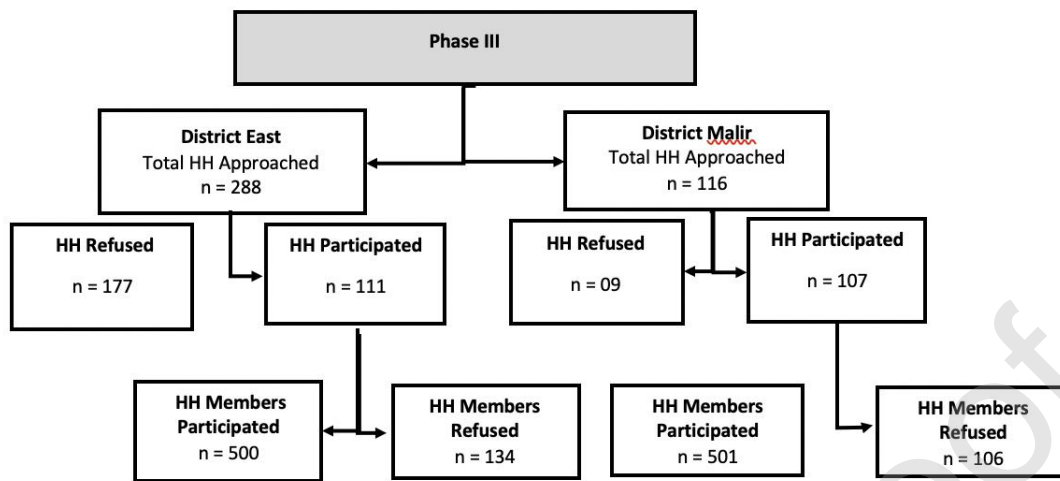


Figure 3. Comparison of COVID-19 positive cases trend from the District East and study sampling sites with an estimation of seroprevalence in Phases 1, 2 and 3.

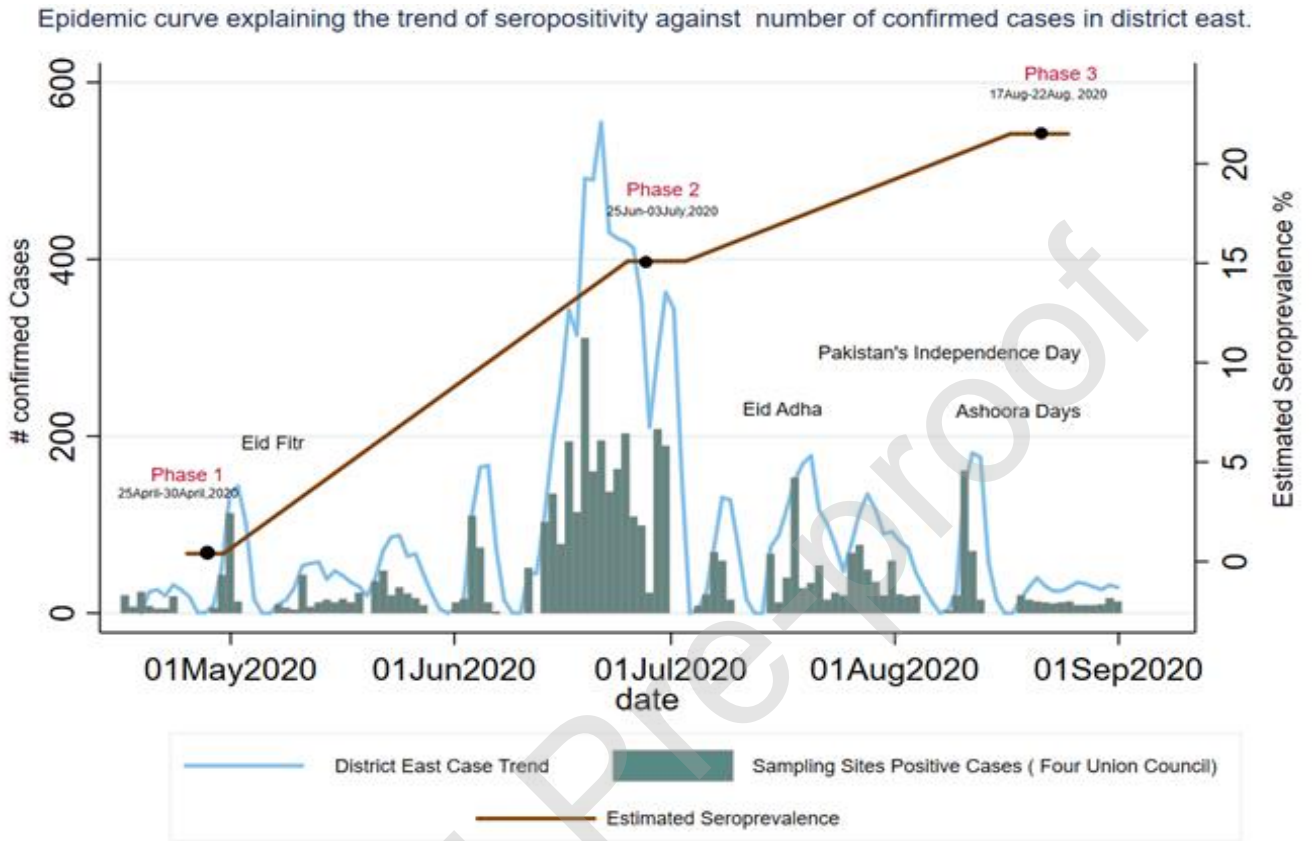


Table 1. General characteristics of the study participants

Characteristics	Districts					
	District East			District Malir		
	Phase 1 (n= 500)	Phase 2 (n= 500)	Phase 3 (n= 500)	Phase 1 (n= 500)	Phase 2 (n= 504)	Phase 3 (n= 501)
Sex, male n (%)	256 (51.2%)	225 (45.0%)	211(42.2%)	207 (41.4%)	199 (39.5%)	206(41.1%)
Age in years , mean(SD)	26.2 (17.9%)	25.9(16.7%)	27.1(17.7%)	28.5 (17.9%)	24.32 (16.7%)	26(16.7%)
Age, categories, n (%)						
0-4 year	35 (7.0%)	22 (4.4%)	31(6.2%)	26 (5.2%)	33 (6.6%)	29(5.8%)
5-9 year	57 (11.4%)	54 (10.8%)	56(11.2%)	52 (10.4%)	74 (14.7%)	75(15.0%)
10-18 year	107 (21.4%)	139 (27.8%)	91(18.2%)	86 (17.2%)	120 (23.9%)	87(17.4%)
19-39 year	185 (37.0%)	170 (34.0%)	206(41.2%)	203 (40.6%)	186 (37.0%)	195(38.9%)

40-59 year	83 (16.6%)	97 (19.4%)	84(16.8%)	103 (20.6%)	65 (12.9%)	90(18.0%)
60 and above	33 (6.6%)	18 (3.6%)	32(6.4%)	30 (6.0%)	25 (5.0%)	25(5.0%)
Household size, mean (SD)	6.1(4.17)	6.6(3.6)	5.7(2.3)	6(2.8)	5.89(3.2)	5.6(2.6)
Working outside home, n (%)	155 (31.0%)	145 (29.0%)	107(21.4%)	143 (28.6%)	124 (24.6%)	116(23.2)
Comorbid, n (%)						
None	453 (92.3%)	463 (93.0%)	471(94.2%)	473 (94.6%)	479 (95.0%)	482(96.2%)
Diabetes	20 (4.0%)	14 (2.8%)	8(1.6%)	6 (1.2%)	10 (2.0%)	5(1.0%)
Hypertension	14 (2.8%)	13 (2.6%)	16(3.2%)	16 (3.2%)	11 (2.2%)	14(2.8%)
Asthma or Allergy	3 (0.6%)	3 (0.6%)	3(0.6%)	1 (0.2%)	0 (0.0%)	0(0.0%)
Chronic Hepatitis	2 (0.4%)	3 (0.6%)	1(0.2%)	3 (0.6%)	3 (0.6%)	0(0.0%)
Chronic Heart Disease	0 (0.0%)	3 (0.6%)	1(0.2%)	1 (0.2%)	1 (0.2%)	1(0.2%)
Asymptomatic	435 (87.0%)	453 (90.6%)	471(94.2%)	472 (94.4%)	487 (96.6%)	473(94.4%)
Sought care	9 (1.8%)	9 (1.8%)	2(0.4%)	3 (0.6%)	4 (0.8%)	0(0.0%)

Hospitalization	1 (0.2%)	5 (1.0%)	0(0%)	1 (0.2%)	0 (0.0%)	0(0.0%)
History of Travel	2 (0.4%)	11 (2.2%)	7(1.4%)	8 (1.6%)	1 (0.2%)	4(0.8%)
Contact with suspected or confirmed COVID-19 case	1 (0.2%)	1 (0.2%)	2(0.4%)	0 (0.0%)	1 (0.2%)	1(0.2%)

Journal Pre-proof

Table 2: Age and Gender based prevalence of COVID-19 infection in District East and District Malir for Phase 2 and 3.

Variables		Phase 2		Phase 3	
Gender	Age	District East	District Malir	District East	District Malir
Female	0-4	0.14 (0.04, 0.25)	0.08 (0.02, 0.14)	0.20(0.09, 0.30)	0.13(0.07, 0.21)
Female	5-9	0.13 (0.04, 0.22)	0.09 (0.04, 0.15)	0.20(0.10, 0.23)	0.13(0.07, 0.19)
Female	10-18	0.13 (0.05, 0.21)	0.09 (0.04, 0.14)	0.20(0.12, 0.23)	0.12(0.07, 0.19)
Female	19-39	0.15 (0.07, 0.23)	0.10 (0.05, 0.16)	0.24 (0.16, 0.33)	0.13(0.08, 0.19)
Female	40-59	0.15 (0.07, 0.26)	0.08 (0.03, 0.14)	0.23(0.15, 0.33)	0.14(0.08, 0.21)
Female	60+	0.15 (0.05, 0.27)	0.01 (0.04, 0.19)	0.21(0.12, 0.31)	0.13(0.06, 0.19)
Male	0-4	0.14 (0.04, 0.25)	0.08 (0.02, 0.14)	0.19(0.09, 0.29)	0.12(0.06, 0.19)
Male	5-9	0.13 (0.03, 0.23)	0.08 (0.03, 0.14)	0.19(0.08, 0.28)	0.12(0.06, 0.19)
Male	10-18	0.17 (0.09, 0.28)	0.01 (0.04, 0.17)	0.22(0.14, 0.32)	0.14(0.08, 0.22)
Male	19-39	0.15 (0.08, 0.24)	0.08 (0.03, 0.13)	0.22(0.14, 0.31)	0.12(0.07, 0.18)
Male	40-59	0.22 (0.10, 0.40)	0.08 (0.03, 0.14)	0.22(0.14, 0.35)	0.13(0.08, 0.22)
Male	60+	0.14 (0.05, 0.26)	0.09 (0.03, 0.15)	0.21(0.11, 0.32)	0.13(0.07, 0.20)