medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Running Title: COVID-19 serosurvey in Karachi, Pakistan

Title: Serial household serosurvey for COVID-19 in low and high transmission neighborhoods of urban Pakistan.

Authors: Muhammad Imran Nisar MBBS, MSc; Nadia Ansari MBBS, MSc; Mashal Amin BDS, MSc; Aneeta Hotwani MSc; Farah Khalid MSc; Najeeb Rehman BSc; Arjumand Rizvi MPhil; Arslan Memon MBBS, MPH; Zahoor Ahmed MBBS; Ashfaque Ahmed MBBS; Junaid Iqbal PhD; Ali Faisal Saleem MBBS, MSc; Uzma Bashir Aamir MBBS, PhD; Daniel B Larremore PhD; Bailey Fosdick PhD; Fyezah Jehan MBBS, MSc

Affiliations:

The Aga Khan University, Karachi, Sind, Pakistan (MI. Nisar, N. Ansari, M. Amin, A. Hotwani, F. Khalid, N. Rahman, A. Rizvi, J. Iqbal, AF. Saleem, F. Jehan)

Health Department, Government of Sind, Karachi, Pakistan (A. Memon, Z. Ahmed, A. Ahmed)

World Health Organization Country Office, Karachi, Sind, Pakistan (UB. Aamir)

University of Colorado, Boulder, Colorado, USA (DB. Larremore)

Colorado State University, Fort Collins, Colorado, USA (B. Fosdick)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract- 150 words

Serial household antibody sero-surveys informs the pandemic where testing is non-uniform. Young populations with intergenerational co-residence may have different transmission dynamics. We conducted two serial cross-sectional surveys in April and June 2020 in low- and high-transmission neighborhoods of Karachi, Pakistan, using random sampling. Symptoms were assessed and blood tested for antibody using chemiluminescence. Seroprevalence was adjusted using Bayesian regression and post stratification. CRI with 95% confidence intervals was obtained. We enrolled 2004 participants from 406 households. In June 8.7% (95% CI 5.1-13.1) and 15.1% (95% CI 9.4 -21.7) were infected in low- and high-transmission-areas respectively compared with 0.2% (95% CI 0-0.7) and 0.4% (95% CI 0 - 1.3) in April. Conditional risk of infection was 0.31 (95% CI 0.16-0.47) and 0.41(95% CI 0.28-0.52) in District Malir & District East respectively with overall only 5.4% symptomatic. Rapid increase in seroprevalence from baseline is seen in Karachi, with a high probability of infection within household.

Keywords: Seroepidemiologic Studies; Coronavirus; Pakistan

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Text-2541 words

INTRODUCTION

The global COVID-19 pandemic has resulted in more than 27 million confirmed cases (until September 07) and more than 883,000 deaths, with an estimated case fatality rate (CFR) of 3.27% % [1]. Pakistan was among the first of low- and middle- income countries (LMICs) to be affected, and since then, there have been 298,903 cases with 6345 deaths (CFR 2.1 %) [2]. Karachi, a large metropolitan city became the epicenter of epidemic on 26th Feb 2020. Since then, Karachi has seen the largest number of cases (~83000 or 31% of all cases) in Pakistan.

The population demographics of Pakistan are typical of most LMICs with over 65% younger than 30 years of age, nuclear families with intergenerational co-residence and average household sizes of 6 or greater. Karachi, has been a hotbed for COVID-19 infections in the country. Crowded neighborhoods, urban slum dwellings and poor adherence to social distancing measures add to the dynamics of infection transmission. Similar to other countries, the pandemic is monitored through symptom-based surveillance, but supply (equipment, reagents, and nasopharyngeal-swabs) and demand side issues (stigma and fear) limit its effectiveness along with undetected pre-symptomatic and asymptomatic transmission [3].

In this milieu, population-based sero surveys can be ideal [4, 5] however high quality population based surveys may not be feasible in an ongoing pandemic because they mandate training on personal protective equipment (PPE) for data collectors and allaying fear or anxiety in potential participants to avoid non-response bias. Sampling entire households can ease

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

procedures of data and blood collection and provide an opportunity to study transmission at household level especially when lockdown and social distancing measures are in place.

Household transmission is a concern in closed congested neighborhoods of metropolitan cities when lockdown measures are in place. It is also of importance because of intergenerational co-residence of young and old, especially in light of recent evidence indicating that households with individuals >60 years of age are at risk of more severe disease[6]. 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%) [7]. However the exercise is resource intensive and transmission may differ between symptomatic and asymptomatic households, as symptomatic individuals are more likely to transmit the virus [8-11]. In a seroprevalence data where household transmission cannot be determined, 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.

With this approach, we estimated changes in seroprevalence in low- and high-transmission neighborhoods of Karachi between April and June 2020 and compared the seroprevalence to total reported positive tests in the area to understand the adequacy of reporting from NP-PCR testing. We also determined age - gender stratified estimates of seroprevalence and assessed the conditional risk of infection (CRI) by adapting the World Health Organization (WHO) Unity protocol[12].

MATERIALS and METHODS

Study Participants and sample collection

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

We conducted the study in two areas. Four sub-administrative units (union councils) of District East were purposely selected as high-transmission area based on highest number of cases to date. One union council of District Malir was selected, where case reporting had been poor and only 164 NP-PCR tests had been done until June 2020 of which 4 were postitive, was designated as a low-transmission area (Figure 1).

Two serial cross-sectional surveys were performed at the household level, between April 15-25 and June 25- July 11, 2020. Four teams, comprising of one data collector and one phlebotomist each conducted the survey. A detailed list of cases was available in District East which allowed households to be selected through systematic random sampling as follows: Reference point was identified randomly from the list of current positive cases and spinning a bottle or pen indicated direction. The reference household with the positive case did not take part in the survey, instead every nth household was sampled based on second to last digit of a bank note until the lane ended in which case the team moved to the next lane or until the day was completed. Same sampling strategy was repeated daily until the overall sample size of the survey was achieved. In case of refusal from a household, the household on the right was sampled. Household level approval and individual participant written informed consent or assent was obtained. A slightly different approach was adapted in District Malir where a list of active cases was not available but a line listing of households from demographic surveillance [13] allowed simple random sampling to be used for household identification daily until the sample size was achieved. In both areas, all household members were eligible to participate irrespective of their infection status.

All team members were trained in use of PPE, hand hygiene, disinfection techniques and safe transportation of biological samples. A 3 ml sample of blood from infants and 5 ml from

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

those older than 1 year was collected by a trained phlebotomist and transported to the Infectious Diseases Research Lab for centrifugation, serum separation and storage at -20^{0}C. Demographic and clinical data were collected using the adapted survey instrument from the WHO Unity Protocol for sero-epidemiology survey of COVID-19 [4].

Laboratory analysis

A commercial Elecsys® Anti-SARS-CoV-2 immunoassay (Roche Diagnostics), targeting both IgG and IgM against SARS-CoV-2 was performed at the Nutritional Research Laboratory (NRL) at Aga Khan University. The manufacturer reports 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 post a PCR positive test [14]. The assay was optimized at the NRL on 20 stored sera from before 2019, and 25 newly collected sera from RT-PCR-confirmed COVID-19 cases. All 20 pre-pandemic sera were negative while 5 of 25 sera from cases were negative.

STATISTICAL ANALYSIS

Sample Size

The sample size for each phase of the survey per site was 500, totalling 1000 participants per survey. This allowed us to estimate an age adjusted prevalence for each site from 20-30 % at 95 % confidence with precision of ± 5% and a design effect of 1.5 for household level clustering.

Data Management

All data was double entered on an SQL database and checked for completeness and consistency. Information was collected on age, gender and household size was collected along with occupation, history of travel and contact. Reported comorbidities, presence of symptoms,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

care seeking and hospitalization were recorded. Age was presented as mean and standard deviations (SD) and also categorized. For symptoms, history was taken regarding presence of fever, respiratory symptoms such as sore throat, shortness of breath, chest pain etc. in previous two months (Supplemental Appendix 3). Occupation was recoded to create a category of those working outside home and not. Those reporting fever or respiratory symptoms in the last two months were categorized as symptomatic and presented as proportions.

Estimation of overall, age and gender stratified seroprevalence

Age and gender-stratified seroprevalence estimates were computed using a Bayesian hierarchical regression model. This approach, described in detail in the Supplementary Material, accounts for uncertainty due to finite lab validation data,[15] and produces estimates with uncertainty across age and gender groups using typical choices of uninformative or weakly informative prior distributions[16, 17].

Seroprevalence estimates were computed for each district and each survey phase independently, and all seroprevalence estimates are expressed as posterior means and 95% equal-tailed credible intervals based on 20000 samples from the Bayesian posterior distribution. All calculations were performed in R and samples from the posterior distributions were obtained using Stan [18].

Household conditional risk of infection (CRI) analysis

CRI, the probability that an individual in a household is infected, given that another household member is infected[19] was estimated. CRI is presented as a fraction whose 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

first individual in the pair is infected. A 95% confidence interval was estimated via bootstrap for each area by resampling households with replacement.

Results

A total of 2004 participants were enrolled across two phases from District East and District Malir. Figure 2 Panel A and B describe the flow of participants. There were high refusal rates in both the areas at the household level, 68% and 43% in district East and 44% and 42% in district Malir participated in phase one and two respectively. The GIS location of refused and accepted households is presented in supplement figure in appendix. Among households who agreed to take part, individual participation rate was 82.3% (1000 out of 1215 household members eligible) in phase 1 and 76.5% (1004 out of 1312 household members eligible) in phase 2. Table 2 describes the baseline demographic and clinical characteristics of the enrolled participants.

In Phase 1 of the study, only 2 of 500 samples tested positive in District East and 0 of 500 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. Both districts showed marked and significant increase in seroprevalence between time points. Overall reporting of symptoms was low, 5.7% only. Of the total 166 participants who tested positive, only 9 (10.2 %) gave a history of fever or respiratory symptoms or both in the last 2 months (Supplemental Table)

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 phase two. CRI estimates were 0.41 (95% CI 0.28-0.52) in District East and 0.31 (95% CI 0.16-0.4) in District Malir.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Given the correlation among household seropositivity values, we estimated seroprevalence via a Bayesian regression model (see Methods) which took into account household membership, age, and gender for each individual. We adjusted for test accuracy by modeling directly on the the lab validation data reported by the test manufacturer[14]. Seroprevalence estimates by age and gender were then post stratified to adjust for the demographic makeup of the respective district.

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, post-stratified seroprevalence 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. The rise in seroprevalence in district East corresponded with the epidemiology as ascertained through daily case reporting in the district as well as the 4 study sampling sites. (Figure 2). Seropositivity rates were indistinguishable between male and female within each district as well as between age groups as seen in Figure 3.

Discussion

The two serial serosurveys conducted in in low- and high- transmission neighborhoods of the largest metropolitan city of Pakistan indicate a rapid increase in seroprevalence rates between April and June. We are unable to comment on the differences between the two neighborhoods based on the differential sampling methodologies but have shown an independent steep rise in seroprevalence in both high and low transmission neighborhoods of Karachi, correlating with the rise of the epidemic in District East (Figure 4). Data for district Malir was not available for comparison but verbal reports shared by the District Health Office indicated that there were only 4 positive cases out of only 164 tested until the period of the second survey. The

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

low number of tests as well as the positivity are likely a result of incomplete reporting and surveillance.

The survey did not identify any difference in seroprevalence between males and females or age categories. Although prevalence appeared to be increasing with age with the greatest probability at extremes of age, the 95% confidence intervals overlapped. Age related prevalence of SARS-CoV2 has been variably reported[20] but the age patterns seen in our study are fairly consistent with other studies[21]. While seroprevalence appeared to be similar, age and gender related reporting of cases as demonstrated in District East, mortality was much higher (Figure 4) showing that the illness is more severe in elderly males [22-24].

To our knowledge, our study is the first published seroprevalence study in Pakistan and South Asia, with available baseline rates of seroprevalence for comparison. Other national level serosurveys in Pakistan and other South Asian countries are ongoing. Preliminary reports of seroprevalence from other metropolitan areas in low- middle- income neighborhoods in Mumbai and Pune have shown very high rates of seroprevalence, up to 55% with cases and deaths at a high[25], while a survey carried out in Guilan province, Iran showed a prevalence of 33% at the height of its epidemic phase [24, 26]. In contrast, the epidemic in Karachi appears to be on a decline even at the relatively low seroprevalence rates seen in our study. While presence of antibodies cannot be equated to presence of protection, this differential in seroprevalence in context with the stage of pandemic in the three cities attributes to some uncertainty about the threshold for herd immunity[27]. It also requires that a closer look be taken at population demographics, dwelling type (concrete or other) and slum setting (urban or periurban) when comparing the geographical differences in COVID-19 cases, deaths and incidence. The first phase of the study was done early on in the pandemic 3 weeks after a provincial lockdown. A

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

higher proportion of the participants reported to be working outside home during this phase as compared to phase 2, however there was no difference in seropositivity seen between those working outside home from those not working. The second phase was conducted after lockdown measures were eased in anticipation of the religious festival of Eid. This reaffirms the efficacy of lockdown measures especially when implemented at the beginning of a pandemic when a novel infection introduced into a naïve population is highly transmissible.

Our survey found a large number of asymptomatic sero-positives, only 1 out of 10 reported any respiratory symptoms, with or without fever. The proportion of asymptomatic infections is reported to be much lower, 27.7% (95% CI, 16.4%□42.7%) in published meta analysis[28]. However high rates of asymptomatic infection have also been reported in India. As per WHO and the Indian Council of Medical Research (ICMR), India, the asymptomatic cases appear to be about 80% compared to the 20% that are symptomatic [29]. Low symptomaticity may be related to fear of disclosure or can be genuine and can contextualize why healthcare resources were not inundated in Pakistan.

The elevation in seroprevalence even in an area of presumably low transmission indicates that seroprevalence studies may serve as effective and low cost tools to determine spread of infection in populations where disease is mostly asymptomatic, ascertainment is low due to fear of diagnostic testing or where access to testing is poor[30]. Monitoring such populations through serial serosurveys can detect resurgence especially when lockdown is eased and scarce resources are moved away from active surveillance and contact tracing. Heterogeneity between low and high income neighborhoods is likely and has been suggested previously [31-33]. This cannot be conceded from our study because we had a different sampling strategy in the two neighborhoods. Census data for households in order to do simple random sampling was available only for

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

District Malir. In absence of such data, we employed systematic random sampling in District around a reference household, reporting a case in the last 2 months. This may overestimate absolute seroprevalence in District East but is still useful to see a temporal comparison of seroprevalence within the Districts independently.

Our results also confirm that close contact within households have a high probability of being infected and should be an important consideration in SARS-CoV2 transmission, especially in areas where there is lockdown and large families are roomed-in in small poorly ventilated spaces, typical of the neighborhoods of Karachi [34]. The probability of an individual to have an infection in the presence of another infected household member in our study, as measured by conditional risk of infection (CRI), was high between 35% - 40%. Secondary attack rate, which is a more reliable indicator of within household transmission, is reported to be lower, 18.8% (95% CI 15.4%-22.2%) in an unpublished review [35]. CRI can function as a substitute in situations where comprehensive surveillance and disease notification strategy is absent and secondary attack rates are difficult to estimate.

The strength of our study is knowledge of baseline seropositivity absent in most serosurveys and the serial nature with a two month inter survey interval allowing for a study of change in seroprevalence. A third round of survey was done in August 2020. About one third or more of our sample included children less than 18 years of age focusing on an understudied age group in the pandemic. Adaptation of the UNITY protocol would allow for pooling of our results with other reports in future. The Electro Chemi Luminescence (ECL) technology used to test antibodies is sensitive and precise as per manufactures report [14].

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Our study had several limitations. The geographical area of the study was limited and the sampling strategy different, not allowing for comparisons between geographies. We had high rates of household level refusal given that the study was conducted in the midst of a pandemic when sentiments of fear and stigma were at a high. We did not do an in-house validation on local samples due to a limited supply chain of testing kits in Pakistan, however this was somewhat compensated by modeling directly on the data reported by the manufacturer.

Conclusion

There is a rapid increase in seroprevalence to SARS-CoV-2 even in areas where transmission is reportedly low. Most seropositives are reported to be asymptomatic and a majority of the population is still seronegative. There is high probability of an individual to be infected given exposure to another infected in the household, irrespective of symptoms. Enhanced surveillance activities of COVID-19 are required especially in low-transmission sites in order to determine the real direction of the pandemic and the risks of household transmission in tightly knit neighborhoods in urban LMIC settings.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

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

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.

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.

Address for correspondence: Fyezah Jehan, Associate Professor, Department of Paediatrics and Child Health, Aga Khan University, Stadium Road, Karachi 74800, Pakistan; Email: Fyezah.jehan@aku.edu; Tel: +92 21 3486 4981

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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. University JH. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Available at: https://coronavirus.jhu.edu/map.html Accessed 24/7/2020 2020.
2. Pakistan Go. COVID-19 Health Advisory Platform by Ministry of National Health Services Regulations and Coordination. Available at: https://covid.gov.pk/stats/pakistan. Accessed 24/7/2020 2020.
3. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet infectious diseases **2020**.
4. Organization WH. Population-based age-stratified seroepidemiological investigation protocol for coronavirus 2019 (COVID-19) infection, 26 May 2020: World Health Organization, **2020**.
5. Winter AK, Martinez ME, Cutts FT, et al. Benefits and challenges in using seroprevalence data to inform models for measles and rubella elimination. The Journal of infectious diseases **2018**; 218:355-64.
6. Aparicio A GS. Intergenerational Residence Patterns and COVID-19 Fatalities in the EU and the US. DISCUSSION PAPER SERIES. Bonn. Germany: IZA – Institute of Labor Economics, **2020**.
7. Wu J, Huang Y, Tu C, et al. Household Transmission of SARS-CoV-2, Zhuhai, China, 2020. Clinical Infectious Diseases **2020**.
8. Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. Nature **2020**:1-.
9. Gudbjartsson DF, Helgason A, Jonsson H, et al. Spread of SARS-CoV-2 in the Icelandic population. New England Journal of Medicine **2020**.
10. Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science **2020**; 368.
11. Streeck H, Schulte B, Kuemmerer B, et al. Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. medrxiv **2020**.
12. Organization WH. Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigations Protocols. Available at: Coronavirus disease (COVID-19) technical guidance: The Unity Studies: Early Investigations Protocols. 2020.
13. Ilyas M, Naeem K, Fatima U, et al. Profile: health and demographic surveillance system in peri-urban areas of Karachi, Pakistan. Gates Open Research **2018**; 2:2.
14. Ltd. RDI. ElectroChemiLuminescence (ECL). Available at: https://diagnostics.roche.com/global/en/article-listing/electroChemiLuminescence-unique-immunoassay-technology.html.
15. Larremore DB, Fosdick BK, Zhang S, Grad YH. Jointly modeling prevalence, sensitivity and specificity for optimal sample allocation. bioRxiv **2020**.
16. Larremore DB, Fosdick BK, Bubar KM, et al. Estimating SARS-CoV-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. medRxiv **2020**.
17. Gelman A, Carpenter B. Bayesian analysis of tests with unknown specificity and sensitivity. medRxiv **2020**.
18. Stan - Stan. Available at: https://mc-stan.org/about/. Accessed 25/7/2020 2020.
19. Curmei M, Ilyas A, Evans O, Steinhardt J. Estimating Household Transmission of SARS-CoV-2. medRxiv **2020**.
20. 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**.
21. 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**.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

22. Li H, Wang S, Zhong F, et al. Age-dependent risks of Incidence and Mortality of COVID-19 in Hubei Province and Other Parts of China. Frontiers in Medicine **2020**; 7:190.

23. Clark A, Jit M, Warren-Gash C, et al. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. The Lancet Global Health **2020**; 8:e1003-e17.

24. Shakiba M, Nazari SSH, Mehrabian F, Rezvani SM, Ghasempour Z, Heidarzadeh A. Seroprevalence of COVID-19 virus infection in Guilan province, Iran. medRxiv **2020**.

25. Delhi's sero-prevalence study finds 23.48 per cent people affected by COVID-19. The Times of India. Jul 22, 2020 ed. New Delhi, India, **2020**.

26. 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**.

27. Rubin R. Difficult to Determine Herd Immunity Threshold for COVID-19. JAMA **2020**; 324:732-.

28. 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**.

29. Chatterjee S, Sarkar A, Chatterjee S, Karmakar M, Paul R. Studying the progress of COVID-19 outbreak in India using SIRD model. medRxiv **2020**.

30. Metcalf CJE, Farrar J, Cutts FT, et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. The Lancet **2016**; 388:728-30.

31. Hooper MW, Nápoles AM, Pérez-Stable EJ. COVID-19 and racial/ethnic disparities. Jama **2020**.

32. de Souza WM, Buss LF, da Silva Candido D, et al. Epidemiological and clinical characteristics of the COVID-19 epidemic in Brazil. Nature Human Behaviour **2020**:1-9.

33. 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.

34. Statistics GoPBo. Housing units by number of rooms and type Available at: www.pbs.gov.pk.

35. 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**.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. General characteristics of the study participants

	District East		District Malir	
Characteristics	Phase 1 (n= 500)	Phase 2 (n= 500)	Phase 1 (n= 500)	Phase 2 (n= 504)
Sex, male n (%)	256 (51.2%)	225 (45.0%)	207 (41.4%)	199 (39.5%)
Age in years , mean(SD)	26.2 (17.9)	25.9(16.7)	28.5 (17.9)	24.32 (16.7)
Age, categories, n (%)				
0-4 year	35 (7.0%)	22 (4.4%)	26 (5.2%)	33 (6.6%)
5-9 year	57 (11.4%)	54 (10.8%)	52 (10.4%)	74 (14.7%)
10-18 year	107 (21.4%)	139 (27.8%)	86 (17.2%)	120 (23.9%)
19-39 year	185 (37.0%)	170 (34.0%)	203 (40.6%)	186 (37.0%)
40-59 year	83 (16.6%)	97 (19.4%)	103 (20.6%)	65 (12.9%)
60 and above	33 (6.6%)	18 (3.6%)	30 (6.0%)	25 (5.0%)
Household size, mean (SD)	6.1(4.17)	6.6(3.6)	6(2.8)	5.89(3.2)
Working outside home, n (%)	155 (31.0%)	145 (29.0%)	143 (28.6%)	124 (24.6%)
Comorbid, n (%)				
None	453 (92.3%)	463 (93.0%)	473 (94.6%)	479 (95.0%)
Diabetes	20 (4.0%)	14 (2.8%)	6 (1.2%)	10 (2.0%)
Hypertension	14 (2.8%)	13 (2.6%)	16 (3.2%)	11 (2.2%)
Asthma or Allergy	3 (0.6%)	3 (0.6%)	1 (0.2%)	0 (0.0%)
Chronic Hepatitis	2 (0.4%)	3 (0.6%)	3 (0.6%)	3 (0.6%)
Chronic Heart Disease	0 (0.0%)	3 (0.6%)	1 (0.2%)	1 (0.2%)
Asymptomatic	435 (87.0%)	453 (90.6%)	472 (94.4%)	487 (96.6%)

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Sought care	9 (1.8%)	9 (1.8%)	3 (0.6%)	4 (0.8%)
Hospitalization	1 (0.2%)	5 (1.0%)	1 (0.2%)	0 (0.0%)
History of Travel	2 (0.4%)	11 (2.2%)	8 (1.6%)	1 (0.2%)
Contact with suspected or confirmed COVID-19 case	1 (0.2%)	1 (0.2%)	0 (0.0%)	1 (0.2%)

Figure 1. Study sites.

Figure 2 Panel A. Flow chart of participants in the phase 1 of study

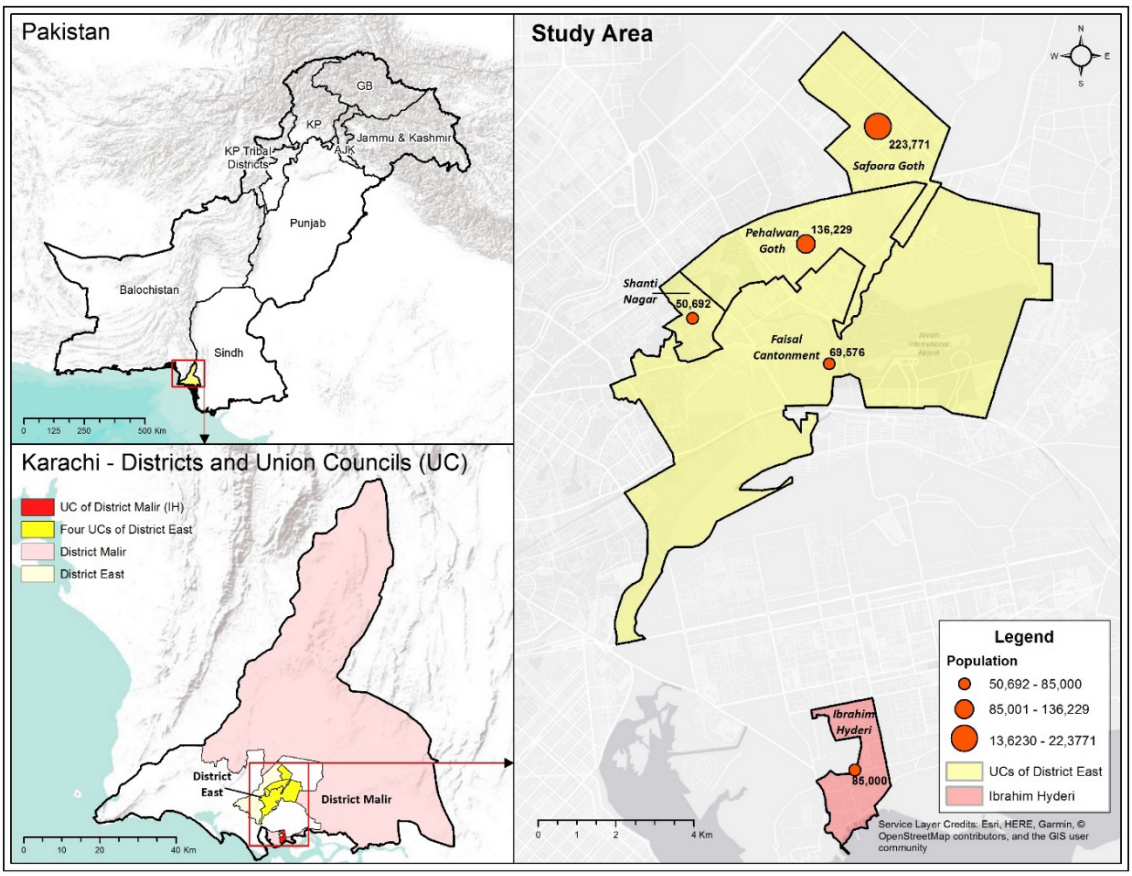
Figure 2 Panel B. Flow chart of participants in the phase two of study

Figure 3: Prevalence estimates by age and gender based on the data from the second survey. The circle represents the posterior mean seroprevalence and the bar represents the 95% equal-tailed credible interval. Posterior mean estimates for District East are consistently greater than those for District Malir, although there is significant overlap in the credible intervals for all age and gender subpopulations. No consistent patterns exist between the prevalence rates for males and females.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Study area and total population



medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

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

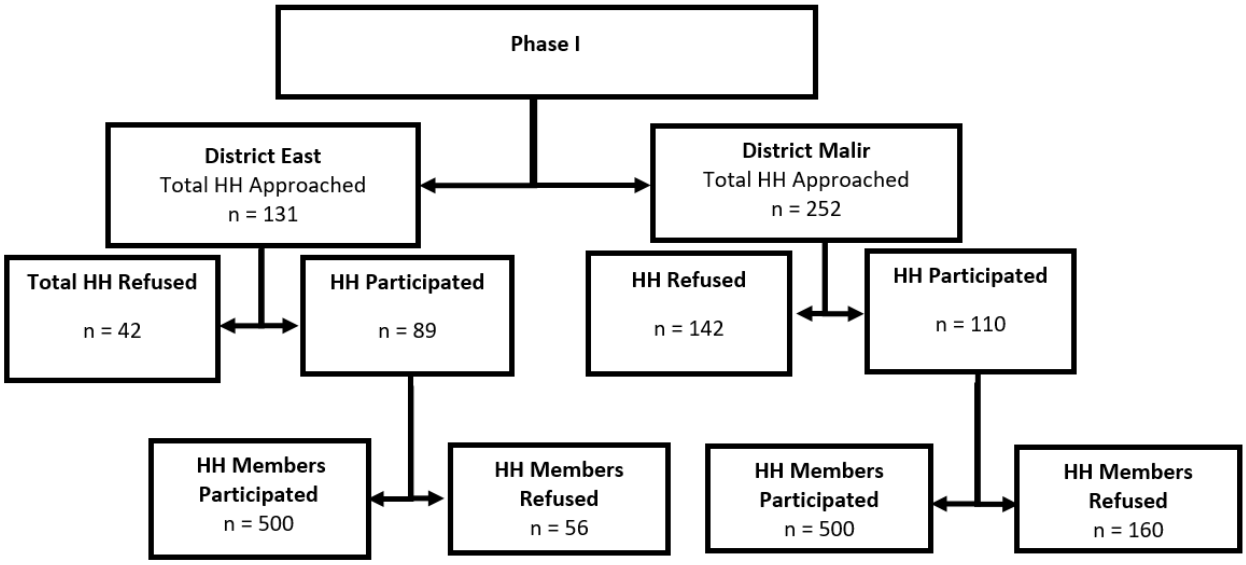
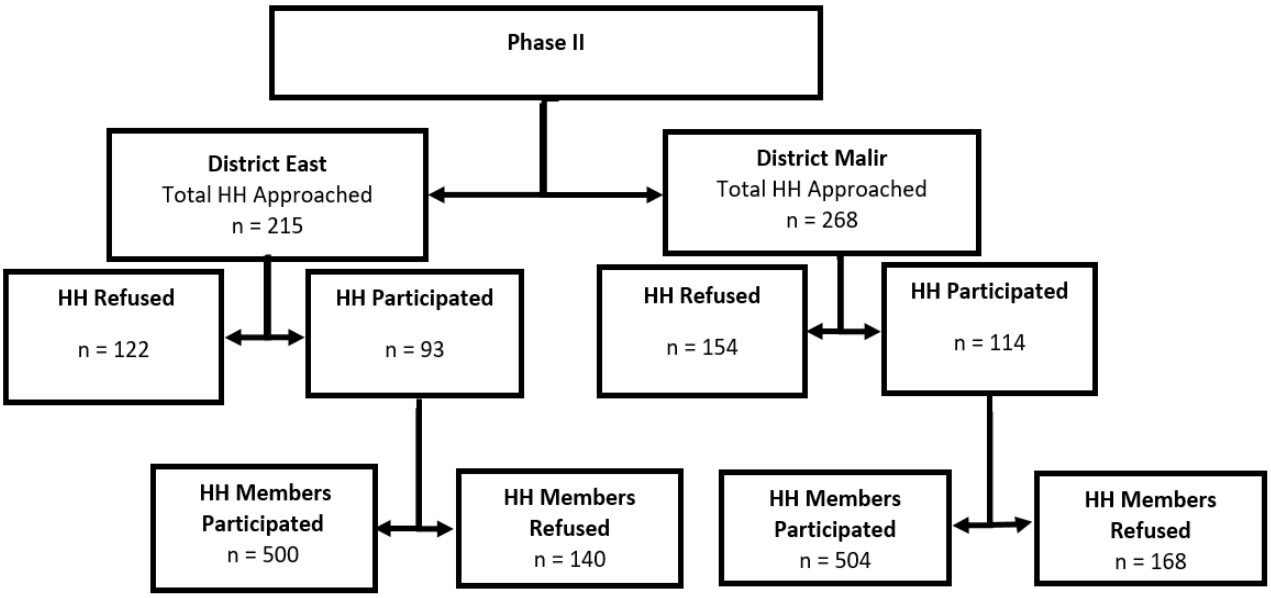


Figure 2 Panel B. Flow chart of participants in phase 2 of study



medRxiv preprint doi: https://doi.org/10.1101/2020.07.28.20163451; this version posted September 22, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY-NC-ND 4.0 International license .

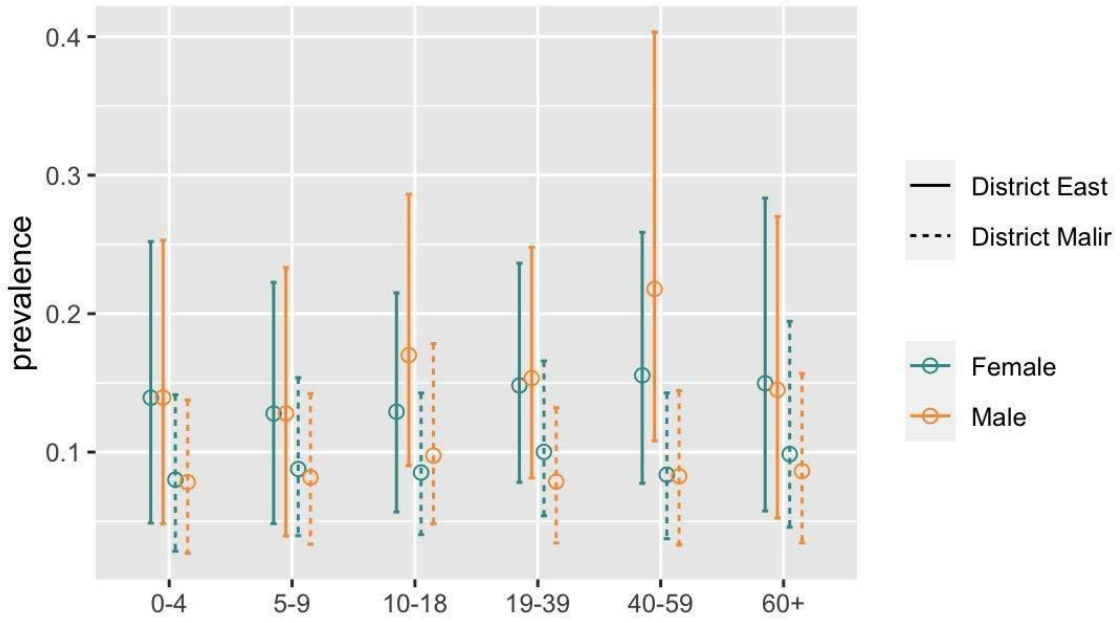



Figure 4: Age and Gender based prevalence of COVID-19 infection in District East and District Malir