

Final report

Smart containment with active learning

Evidence from a
data-driven rapid
policy response

Ahwaz Akhtar
Ali Cheema
Jishnu Das
Adnan Q Khan
Asim I Khwaja
Anum Malkani
Ahmad Mustafa

November 2020

When citing this paper, please
use the title and the following
reference number:
S-20053-PAK-1

IGC

International
Growth Centre



DIRECTED BY



FUNDED BY



Smart Containment with Active Learning: Evidence from a Data-Driven Rapid Policy Response to Covid-19 in Punjab

Ahwaz Akhtar, Ali Cheema, Jishnu Das, Adnan Q Khan, Asim I Khwaja, Anum Malkani, Ahmad Mustafa*

November 2020

Abstract

The problem governments face in designing effective Covid-19 response measures is that there is little data on potential health and economic impacts to figure out tradeoffs between choices and to design a real-time, effective, and evidence-based response. For an effective policy response, it is imperative that governments use data-driven and evidence-based policies that are responsive to on-ground realities, understand the benefits and costs of various policy interventions, and continuously refine policy measures. The Smart Containment with Active Learning (SCALE) approach lays out an evidence-responsive Covid-19 response strategy that seeks to minimize disease spread and limit adverse socio-economic impact by targeting response to the current & projected level of prevalence in the smallest feasible geographical units. Implementing SCALE requires large-scale testing to establish the level of prevalence in each area. The SCALE Smart Testing strategy lays out an effective testing methodology allowing accurate estimates of prevalence given limited testing capacity. The Smart Testing strategy was implemented in Lahore, in collaboration with the Punjab Primary and Secondary Health Department (P&SHD), over a two week period. The strategy was deployed at the census block level and over 12,000 individuals were tested. The findings revealed substantial variation in prevalence rates across census blocks suggesting that the infection spread was localized and a localized policy response may be more effective. Other key findings include: women and the elderly had higher rates of infection; 95% of those who tested positive did not display any symptoms; likelihood of infection increases by pre-existing conditions and the fraction of nearby people who are infected.

*We would also like to thank Captain (retd.) Usman Younis, Secretary Primary & Secondary Healthcare Department and Muhammad Ajmal Bhatti, Special Secretary Primary & Secondary Healthcare Department and Focal Personal of the Government of Punjab for their feedback.

Table of Contents

1. Introduction	3
2. Context	5
3. Methodology	6
4. Results from Analysis of Smart Sampling and Testing Data	7
4.1: Overall Prevalence and Spatial Variation	8
4.2: Prevalence Rates Over Time	10
4.3: Prevalence Rates by Demographic Characteristics	12
4.4: Prevalence Rates for Symptomatic and those with Pre-existing Conditions	15
4.5: What Predicts Covid-19 prevalence? A Multivariate Regression Analysis	17
4.6: Testing Costs and Efficiency.....	20
5. Policy Recommendations	21
6. Conclusion	22
Appendix A: SCALE Operational Plan	23
Appendix B: SCALE Smart Testing Strategy.....	25
Appendix C: Testing and Tracing Survey Instrument.....	30
Bibliography	41

1. Introduction

Governments across the world have grappled with difficult policy questions in response to Covid-19. What has made these policy decisions harder is the surprisingly limited knowledge we had about Covid-19 at the outset. Past experiences, such as the 1918 Spanish Flu outbreak, underscored the importance of early intervention with public health measures such as physical distancing and contact tracing. Early on, countries that adopted such measures, including China and Singapore, saw some success in containing the epidemic (Anderson et al. 2020). However, while strict physical distancing and lockdown measures may mitigate the deaths and medical burden of Covid-19, they could severely dent the economy as well as generate adverse socioeconomic, psychological, and health outcomes.

In recent history, evidence from the influenza pandemics and other outbreaks – for example, H1N1, H5N1, Ebola, Zika – demonstrates the acute economic impact of pandemics and associated public health response measures (Keogh-Brown and Smith 2008, Smith et al. 2009). With economic activity at a stand-still during lockdowns, Covid-19 poses both substantial demand and supply side shocks to the economy. Early evidence suggests that the Covid-19 pandemic and resulting lockdown measures have caused serious economic downturns (Stock 2020, Hausmann and Schetter 2020).

Physical distancing and lockdowns are likely to carry deleterious effects in low-income countries for multiple reasons (Mobarak and Barnett-Howell 2020). Since much of the economy is informal in low-income countries, they struggle to provide financial assistance to those who may need it most. The food chain, from crop production to distribution and sale, is more labor intensive and maintaining food supply may become hard in extended lockdowns. A significant fraction of the disease burden remains acute. Childhood diarrhea, pertussis and Tuberculosis all require regular care, and, in a lockdown, this may not be feasible. Fiscal space at the state/provincial level is limited (Hausmann 2020), and unlike in the U.S. or Europe, where money flows in when times are uncertain, in low-income countries, the money flows out.

Given the potential deleterious impacts of lockdowns, sustained blanket lockdowns for long, drawn out pandemics are not feasible. A growing body of research suggests that containment measures should be based on testing data and/or demographic characteristics, with partial/targeted lockdowns rather than unsustainable blanket shutdowns of the economy (Acemoglu et al. 2020, Eichenbaum, Rebelo and Trabandt 2020, and Alvarez, Argente and Lippi 2020).

This underscores the need for effective testing and data collection strategies. Given the large number of asymptomatic cases, designing an effective testing strategy must go beyond screening for symptoms. And, given limited testing capacity, optimal utilization of existing capacity should leverage innovative solutions such as, smart sampling and pooled testing.

Prevalence and transmission estimates established by testing data will inform not just physical distancing measures but also other key public health policies, including community messaging and enforcement. Appropriate communication and messaging are key for crises in general but especially those where voluntary and informed compliance to policies will have huge implications on the efficacy of policy measures (Nour et al. 2017).

This study addresses the above policy gap by providing a dynamic, targeted, and evidence-based Covid-19 response strategy. The Smart Containment with Active Learning approach utilizes standard frameworks from decision making under policy uncertainty to help governments design better policy responses to the Covid-19 pandemic (Manski 2013). Central to this approach is the Active Learning strategy: using real time testing and refinement of policy responses together with a graded approach that varies by local disease projections.

The four distinct parts of this Active Learning process are:

- i. First, some decisions will essentially remain the same *regardless* of what information becomes available. For instance, the urgent requirement to expand testing capacity and PPE for health workers. In this case, there is no need for information to become available; policymakers can act immediately.
- ii. Second, there are other decisions that are best made after collecting *some* information—especially if that information is relatively costless to collect. These decisions are ones where more information may change the decision. For instance, in countries that imposed sudden lockdowns, a 2-day survey could have helped the government understand that migrants would leave urban areas with a lockdown and this would have allowed multiple mitigation measures to have been put in place. Countries can be prompt in their response without rushing blindly into decisions.
- iii. Third, in making decisions, all prior information should be used. For instance, an understanding of the underlying vulnerability of populations would help governments to undertake a spatially targeted strategy.
- iv. Fourth, every decision will have an impact on the outcomes of interest but will *also* provide further information. This new learning can critically inform the decision-making process tomorrow. A decision that may have little impact on potential outcomes but can massively increase information should be favored. This moves us away from *passive to active learning* and must be a key component of the strategies in both high- and low-income countries.

Based on the Active Learning strategy, the SCALE alert system and operational plan allows for a staged, time-varying response with rapid learning and refinement. The SCALE operational plan (Appendix A) lays out an evidence-responsive Covid-19 response strategy that seeks to minimize disease spread and limit adverse socio-economic impact by targeting response to the current & projected level of prevalence in the smallest feasible geographical units. Implementing SCALE requires large-scale testing to establish the level of prevalence in each area. The SCALE smart sampling and testing strategy (Appendix B) is an effective testing methodology allowing accurate estimates of prevalence in a cost-effective manner with limited testing capacity.

2. Context

Towards the end of February, Pakistan began recording its first few cases of Covid-19. By May 1st, the rise in cases had reached alarming levels, with over 800 cases per day and Punjab, Pakistan's most populous and urbanized province, recording the highest number of cases. The emerging debate around Covid-19 mitigation measures in Pakistan was between two extremes: either locking down to mitigate Covid-19 spread, or remaining open to prevent economic hardship.

Health is a provincial topic in Pakistan with each provincial health department formulating its own health programs and policies. At the federal level, the Ministry of National Health Services Regulation and Coordination functions as a central coordinating entity. This study leveraged a unique partnership with the provincial Government of Punjab (GoPunjab) to examine how Pakistan can deliver an evidence-based rapid policy response at the right level to maximize lives saved and minimize economic costs.

Based on our advisory, the Primary and Secondary Healthcare Department (P&SHD) notified a working group on April 28 2020, composed of our team, the P&SHD and other government departments, with the following TORs:

- Design and share smart sampling, testing and data collection methodologies to help establish prevalence and transmission rates of Covid-19;
- Based on analysis, assist the government to understand immediate health and economic impact;
- Provide analysis of relevant data on Covid-19 related to smart sampling / testing provided by Government of Punjab and report on the findings, where possible, support government to conduct its own analysis;
- Provide technical input to government about Covid-19 policy response.

As members of this working group, we implemented the Smart Testing Strategy during May 2020 in 645 randomly selected census blocks in Lahore. We generated a repeated cross-sectional sample of Covid-19 testing data in Pakistan to examine and understand transmission vectors.

There are two major policy contributions. First, this project recognizes the imperative need for evidence-based policy making that is responsive to on-ground realities. Given the limited feasibility of sustained blanket lockdowns, and the nature of the Covid-19 pandemic with waves of infection, there is a need for context-specific, spatially and temporally varied policy responses which mitigate disease spread while minimizing the negative socioeconomic impact. The main policy question this project examines is how governments may deliver such a dynamic, data-driven, and evidence-based response.

Second, the project seeks to enhance government capacity to gather, analyze and understand data, as well as deploy limited resources as effectively as possible. Through a close collaboration with the Punjab government, this project provides a proof of concept and blueprint that can be

implemented in other districts, provinces and countries. By determining the extent of the remedial social distancing measures through an iterative process based on existing prevalence rates as well as impact on future prevalence, governments can mitigate socio-economic costs and enhance the efficacy of policy measures. Such an approach requires capacity building for governments to gather and analyze data in real time, and rapidly design policy measures based on local conditions. The policy recommendations and lessons learnt have value for many other countries.

3. Methodology

The SCALE smart sampling and testing strategy is generally focused on providing representative prevalence data at the level of the smallest feasible geographical unit (S-grid) within a district where Covid-19 specific response measures (such as testing, quarantine and lockdowns) can be effectively enforced. The ideal S-grid is one that is minimally feasible (200+ households), integrates with census information such as population density, distribution of age, etc., and can be determined using shape files (geo-referenced boundaries). Refer to Appendix 2 for the detailed smart sampling and testing strategy.

The city of Lahore was chosen to begin implementation. Census blocks were established as the S-grid in Lahore. The census blocks selected for inclusion in the smart sample were adjacent to originally infected blocks, i.e., blocks that were discovered to be infected before implementation. Within each selected S-grid, random GPS points were dropped at a ratio of 1 GPS point per 150 households.

The GPS points in each census block were shared with P&SHD, where the District Health Authority's smartphone-equipped field team was trained on administering a patient survey. The survey instrument is included in Appendix C.

To reduce the cost of testing, the SCALE team proposed pooled testing. Recent work has argued this can be a cost-effective means of testing (Majid, Omer and Khwaja 2020). In this case, a pool of up to ten individuals at each GPS point was selected. The pooling procedure is detailed in Appendix B. The left-hand rule was used to determine the first household to be tested. Within each household, the person with the highest contacts and mobility was selected for testing and was subsequently administered the patient survey to record patient contact information, symptoms, comorbidities, number of contacts, and names of nearby contacts who they considered to be highly mobile. Through this 'daisy chaining' process, a pool of up to 10 individuals was built from 1 GPS point. The collected samples were tested through the pooled testing procedure.¹

¹ Pooled testing works by combining a set of individual specimens into a common pool. If the pool tests negative, all individuals are diagnosed as negative. If the pool tests positive, the individuals within the pool shall have to be retested, and the samples processed individually to identify positive individuals from the negative individuals.

Through this innovative approach, which was first implemented in Punjab as part of this smart sampling and testing exercise, the Punjab government was able to decrease costs while simultaneously enhancing testing capacity (Cherif et al. 2020).

This strategy was deployed for a two-week period in May across Lahore, where the testing period includes the period of strict lockdown in the city. Testing took place in 3 stages: Pilot, Wave 1, and Wave 2. The Pilot stage was implemented on May 3rd and 4th 2020 in 11 blocks that were adjacent to originally infected blocks. The field team was trained to follow the daisy chain process. If the chain broke at any point either because a test recipient refused to share contact information, the referred person was not available, or because the referred person refused a test, the field team moved to the adjacent house/property and followed the same set of procedures as before.

Wave 1 implementation spanned from May 5th to 9th 2020 and covered 282 S-grids that were adjacent to originally infected blocks. While largely the same process was followed as in the Pilot stage, anecdotal evidence emerged (through focus groups and discussions with the survey operators from the District Health Authority) of the field team also using their local knowledge and contacts to identify individuals (within the census block) who they deemed to have high mobility.

Wave 2 implementation spanned from 13th May to 16th May 2020 and covered 449 S-grids that were adjacent to originally infected blocks. P&SHD was advised to implement the following three strategies:

- Track A: daisy chaining by contact referrals
- Track B: conducting frontline worker testing (with emphasis on healthcare workers, police, grocery stores)
- Track C: pool testing those individuals identified by the local P&SHD surveyor

Track A was implemented as a part of this study.

4. Results from Analysis of Smart Sampling and Testing Data

A total of 645 blocks have been covered under Smart Sampling in Lahore as of date. Out of 12,251 unique samples collected, 97.7% have received results after submission to the laboratory whereas 2.3% of results were not obtained due to sample rejection by the laboratory.

The table below breaks down the number of samples by test result:

Result	Count	Percentage
Positive	450	3.67%
Negative	11,516	94.0%
Rejected	285	2.33%
TOTAL	12,251	100%

Table 1: Sample Results

4.1: Overall Prevalence and Spatial Variation

The overall Covid-19 prevalence rate in the sample is 3.76% with roughly similar splits between gender - as 4.43% of females and 3.66% of males tested positive, a statistically significant difference at the 10% level.

The blocks in which testing took place were spread all over Lahore. Because our sampling strategy randomly surveyed blocks around infected blocks, this provides some sense of spatial distribution of prevalence. Our results suggest that there was quite a bit of variation in Covid-19 prevalence rates even within a given city/neighbourhood.

First, at the level of blocks we find that the spread of Covid-19 was localized in most cases with 419 out of 645 blocks testing negative for all those sampled. Among the blocks that did contain at least one positive individual (226 blocks), the distribution of positivity rates looked as follows:

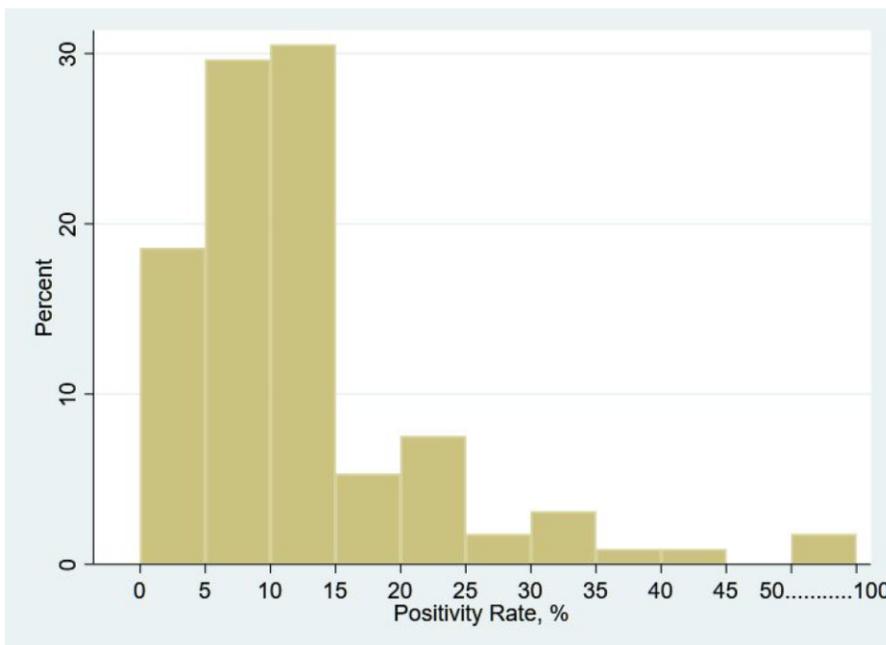


Figure 1: Block-level Prevalence Rates (excluding zero prevalence blocks)

The maps below illustrates this variation spatially and again make the point that there is considerable variation in prevalence rates.

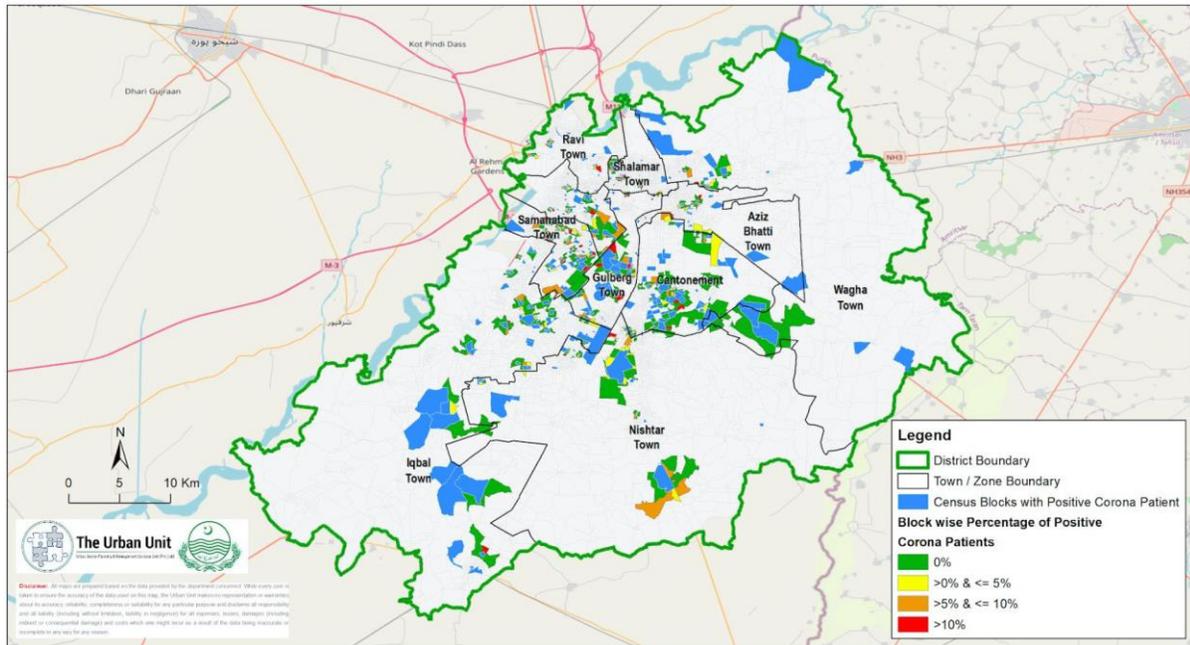


Figure 2: Map of Lahore showing Originally Infected Blocks (in blue) along with Adjacent Blocks where Smart Sampling & Testing was implemented (in green/yellow/orange/red)

Interestingly, there was a lot of very localized variation in prevalence as well. Even adjacent blocks can have fairly different levels of infection suggesting that Covid-19 spread may be a fairly localized phenomenon (especially when lockdown policies are in effect). The figures below provide zoomed-in versions of a few different areas of the overall prevalence map and illustrate the very localized differences in Covid-19 prevalence rates.

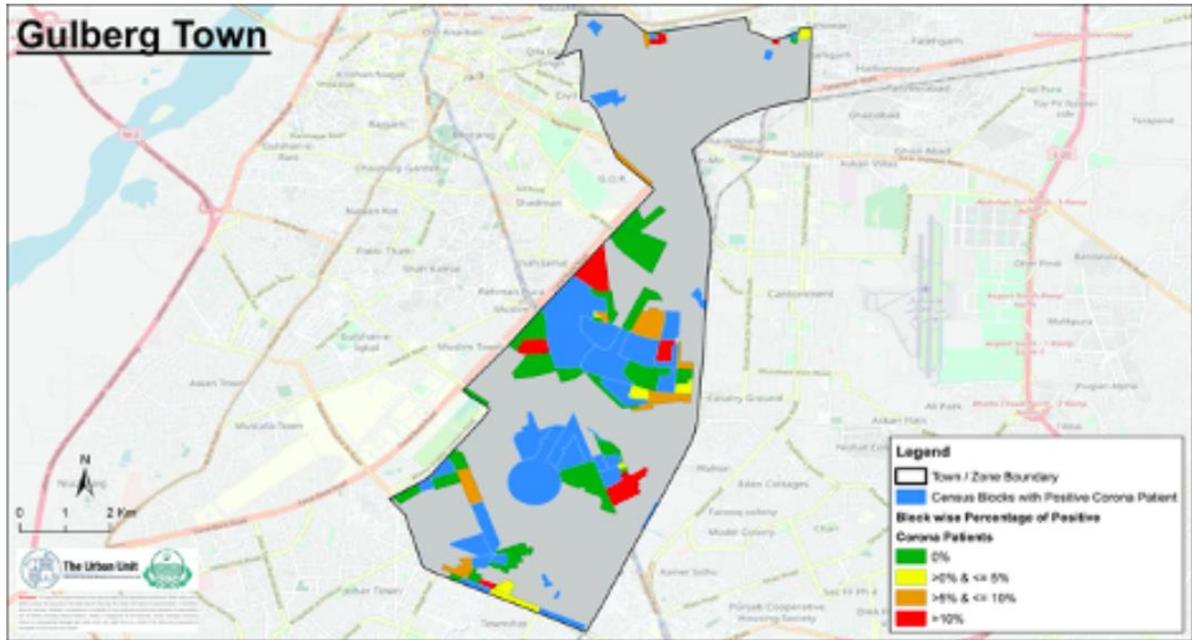


Figure 3: Map of Gulberg Town showing Originally Infected Blocks (in blue) along with Adjacent Blocks where Smart Testing was implemented (in green/yellow/orange/red)

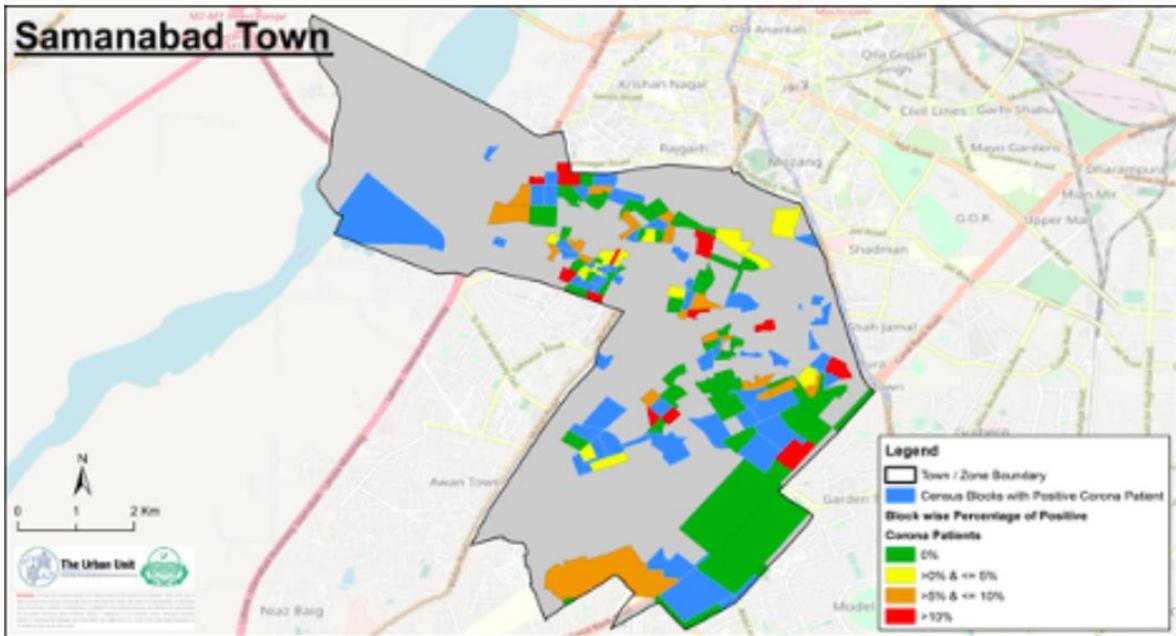


Figure 4 - Map of Samanabad Town showing Originally Infected Blocks (in blue) along with Adjacent Blocks where Smart Testing was implemented (in green/yellow/orange/red)

4.2: Prevalence Rates Over Time

An important statistic that informs how to deal with Covid-19 and the extent and efficacy of lockdowns is how prevalence rates change over time. Without taking any precautionary measures one would expect the number of infected to increase exponentially over time. While we do not have repeat samples for the same area, given that the blocks were selected randomly and tested in no specific order, looking at prevalence rates across the different days of collection provides a rough sense of how they were changing over time.

The figure below shows this evolution from May 5th to 17th 2020. Interestingly, the positivity rate mostly remained between 2.5% and 5% during this period with no evidence of a clear upwards trend let alone an exponential growth.

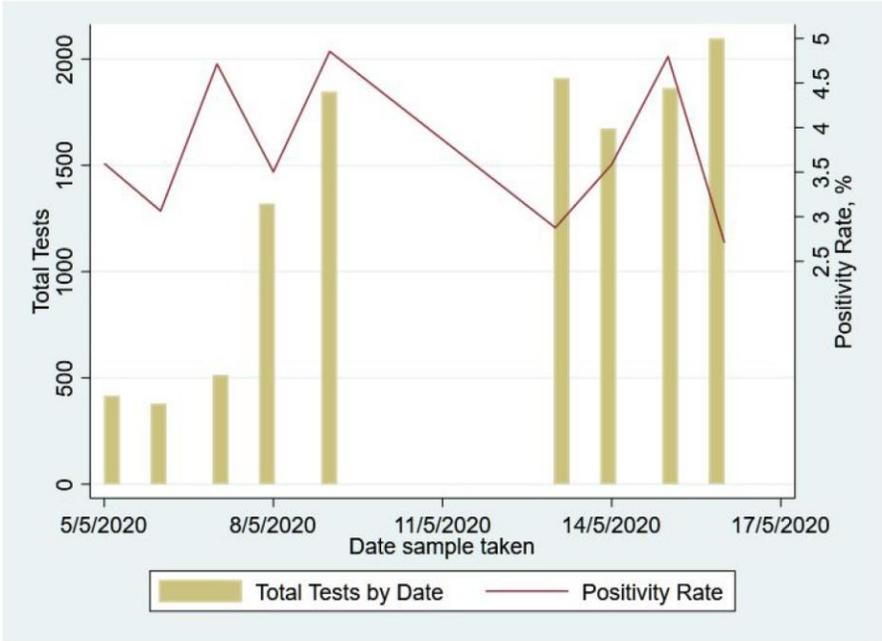


Figure 5: Positivity Rates across Waves 1 and 2

These results, while suggestive, do indicate that the lockdown period seems to have been somewhat effective in the sense that blocks that got surveyed later during this period did not show significantly higher prevalence rates.

However, a more accurate picture of the efficacy of the lockdown measures would have been to have collected repeat samples and to continue testing, especially after the lockdown was eased and, more recently, when a partial lockdown was re-enacted. Unfortunately, despite strong advice given by the SCALE team, this data was not generated and therefore we were unable to accurately assess how well the lockdowns have worked.

One piece of evidence that suggests that initial lockdown was likely to have been effective in mitigating Covid-19 spread, and that the opening up of the lockdown worsened the spread, comes from looking at ventilator usage data (in Lahore) based on statistics provided by Specialized Healthcare & Medication Education Department (SH&ME). We repeat the prevalence

data in the previous figure but now overlay it with time-series data on ventilator usage. Note that while the positivity rate from PCR tests (blue line) provides a closer to real-time measure of Covid-19 prevalence, the number of ventilators used provides an alternative prevalence measure though with greater lag (since those infected may take longer to develop severe enough symptoms that require ventilators).

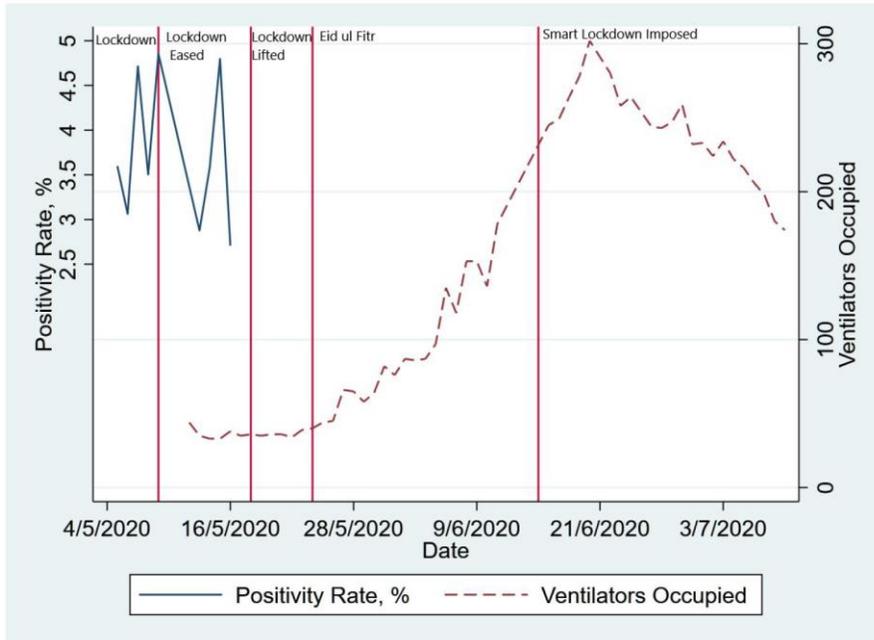


Figure 6: Positivity rate and Ventilator Usage during different policy periods

Nevertheless, comparing these two graphs and using ventilator data to “fill in” the lack of smart PCR testing after mid-May suggests that the initial lockdown may indeed have been effective. First, note that prior to lifting the lockdown both the PCR test-based prevalence data and ventilator usage shows that there was not much change in Covid-19 prevalence. However, we can see that a week after lockdown was lifted, ventilator usage started increasing and continued to increase substantially suggesting that the initial lockdown indeed had been effective and the subsequent opening caused prevalence rates to increase. Further, after the imposition of smart lockdown in mid-June the ventilator usage declined in line with expectations.

4.3: Prevalence Rates by Demographic Characteristics

Age: The median age among all tested was 36 years and therefore the sample skewed younger. In the figure below, the dashed green line represents the (“smoothed out”) relationship between the positivity rate and respondent age.² The bar chart represents the fraction of our sample of a given age.

² We plot the lowest (Locally Weighted Scatterplot Smoothing) curve that shows the smoothed-out relationship between positivity against age.

The main trend to note in the graph is that the rate of Covid-19 positivity increased with age of the individual tested. It is lower for younger age groups and increased at a high rate till age 30, after which it increased but at a lower rate. Among those above 50 years of age (comprising 20% of the sample), the positivity rate exceeded 4% throughout.

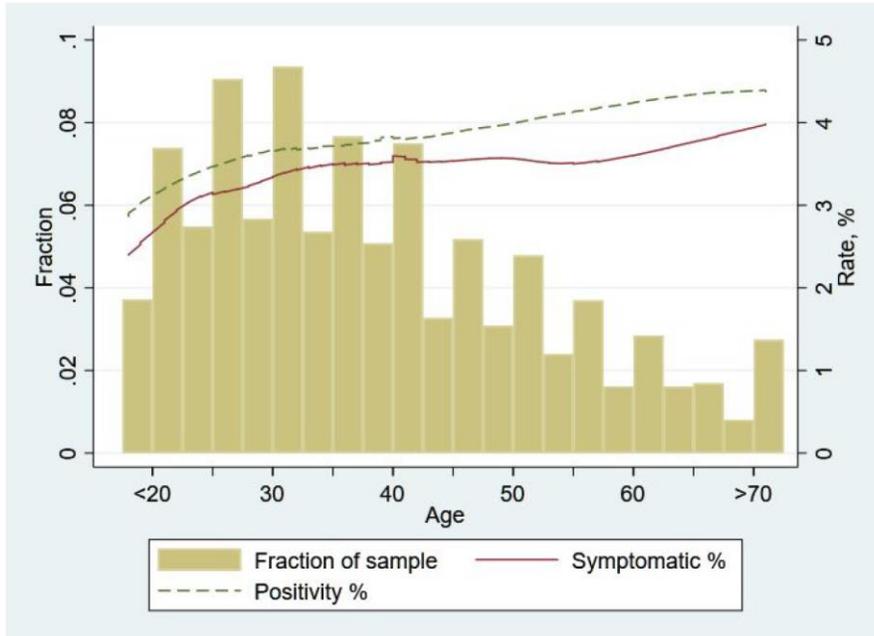


Figure 7: Age Profile, Symptomacy and Positivity Rate of the Sample

We should note that, as shown in the solid red line in the figure, part of this relationship is likely due to the higher fraction of symptomatic people we see at higher age groups. We should caution that since we do not know refusal rates at different age groups we do not know if this is a feature of the underlying population (i.e., older people are more likely to be symptomatic/infected) or is it that older people who are symptomatic are more likely to consent to being tested. In future rounds of testing, refusals and reasons for refusals should be collected to have a better sense of this. Regardless, these results do suggest extra caution for the elderly - not only do we know from other countries experiences that their morbidity outcomes may be worse, but at least based on the data we have it seems they may be more likely to be infected as well.

Gender: Likely due to social norms and the fact that we had suggested that priority be given to testing individuals within the household who had higher contacts/exposure, more men were tested (10,410 valid samples) as compared to women (1,556 valid samples). We had previously noted that the positivity rate seemed significantly higher for women (4.43%) compared to men (3.66%).

Figure 8 below examines this further by examining whether Covid-19 prevalence varies differentially by age for men and women. Interestingly, we see a marked difference here: For men their positivity rate is rising till age 30 but then remains around 4% at all ages beyond. Women, on the other hand, after a small initial drop, show a clear and sustained rise in positivity

rates with age, starting from around 4% even at ages lesser than 20 and then increasing to about 7% at beyond 70 years of age.

Without knowing if there is differential selection in who agrees to be tested it is hard for us to determine what causes this large age effect for women but not for men. While there is some selection here (the figure below shows there is a great fraction of symptomatic women as age increases), it is also not clear that this by itself can drive the large increases women see in being PCR positive as their age increases.

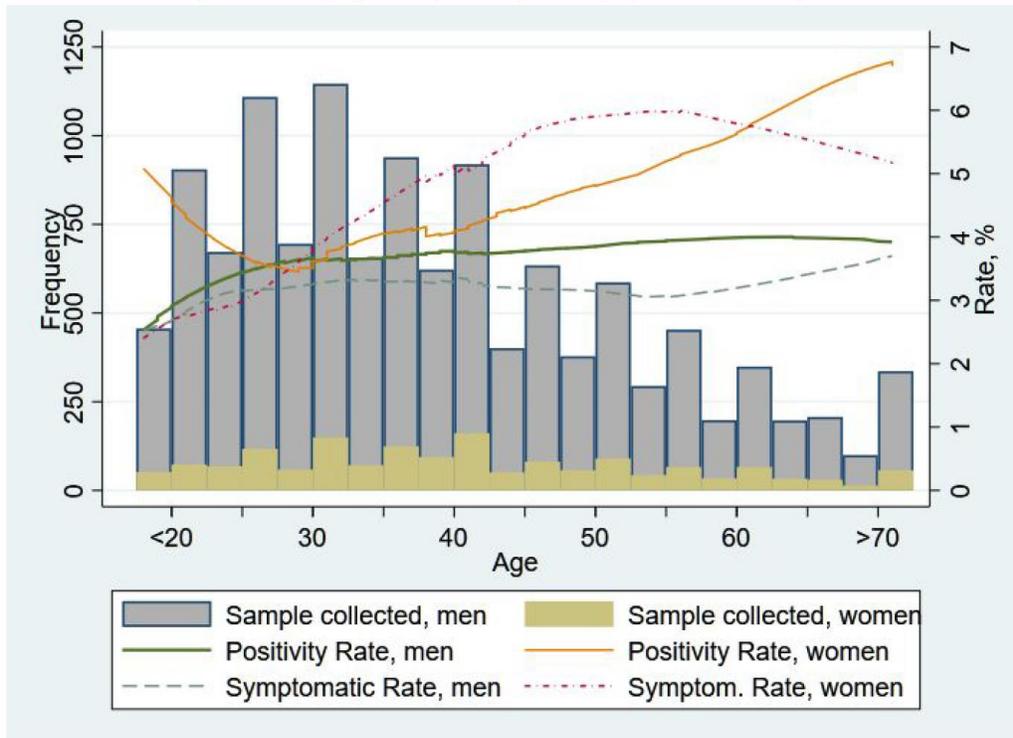


Figure 8: Age Profile, Symptomacy and Positivity Rate, by Sex

Other Attributes: Another risk factor associated with contracting the virus was whether a person had to leave their residence for work. During the sample collection period, the lockdown was eased in Lahore and therefore people might have been able to commute to their workplaces. Those who were employed constituted 85.6% of the sample and showed a generally higher positivity rate as compared to those unemployed across all age groups. Overall positivity rate for those employed stood at 3.8% versus 3.2% for the unemployed, a statistically significant difference at the 10% level. Our data is too limited to draw inferences on whether different occupations have differential risks, but this is something that could be worth examining in future testing.

In the early days of virus spread (February-March 2020), all the cases in Pakistan almost exclusively came from those who had either traveled internationally or their domestic contacts. In our sample, there were 532 cases of travel (4.3%) and among these only 10 had traveled

internationally. In our data, while 4.5% of those who traveled tested positive compared to 3.7% of those who did not, these differences are not statistically significant.

4.4: Prevalence Rates for Symptomatic and those with Pre-existing Conditions

Given PCR testing is costly, one hope is that symptom-based phone surveys may be an effective way of determining Covid-19 prevalence. We therefore examine how effective such methods may be.

In our overall sample, 3.4% of individuals report themselves as having relevant symptoms. While the symptomatic rate is not that far off from the overall rate of Covid-19 prevalence rate and, as we had noted before, symptomatic rate increases with age along similar lines to Covid-19 positivity, the relationship between being Covid-19 positive in a PCR test and reporting symptoms is not as strong as one would have hoped. While symptomatic individuals were more likely to have tested positive in PCR testing (5.7%), this is only slightly greater than the PCR-based Covid-19 prevalence rates in asymptomatic individuals (3.7%). While the difference is statistically significant, with close to 95% of those detected as Covid-19 positive in the PCR tests being asymptomatic, it does not seem symptom based surveys will be an effective way to identify the majority of infected individuals.

Further analysis reveals that for asymptomatic individuals, there is little pattern with age – while there is some initial increase with age, past late-20s the positivity rate remains stable at around 4%. In contrast, for those reporting symptoms, there is (after a small initial dip) a very strong positive relationship between positivity and age with older people exhibiting substantially higher positivity rates. For individuals reporting symptoms, those greater than 50 years of age have a prevalence rate of around 13%, which is significantly higher compared to those less than 50 (3.8%). This suggests that while symptom based identification may be generally less useful, it is more indicative of Covid-19 prevalence for older individuals, i.e., those above 50 years old are more than three times more likely to show Covid-19 positivity if they are symptomatic than if they are not. Though even in this older population around 91% of those detected as positive in PCR-testing report no symptoms.

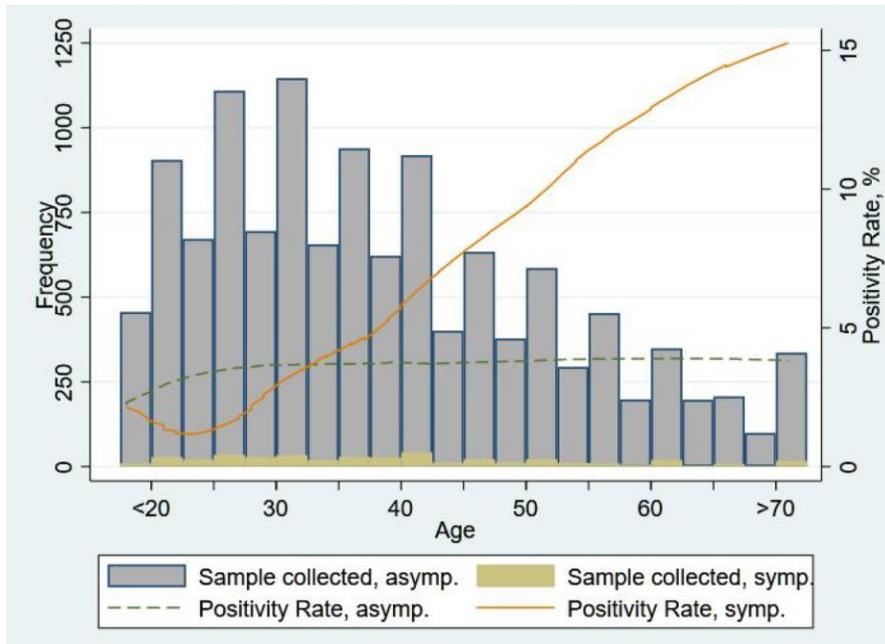


Figure 9: Age Profile and Positivity Rate by Symptomatology

It is also instructive to examine whether individuals with pre-existing conditions show differential rates of positivity. In our sample, 707 (5.8%) people reported having any pre-existing conditions such as diabetes, hypertension, breathing issues, etc. Given what we know about underlying prevalence, we suspect there may be some under-reporting here. Nevertheless, we find that the overall positivity rate of those with pre-existing conditions was 5.8% compared to 3.6% with no pre-existing conditions and this difference is statistically significant. Among those with no conditions, the positivity rate rises in earlier ages and then remains around 4% into higher age brackets. However, those with pre-existing conditions exhibit an increasing positivity rate with age, peaking at around 10% among the elderly. Curiously and somewhat puzzlingly, this drops at above 60 years of age though that may be driven by differential reporting and smaller samples.

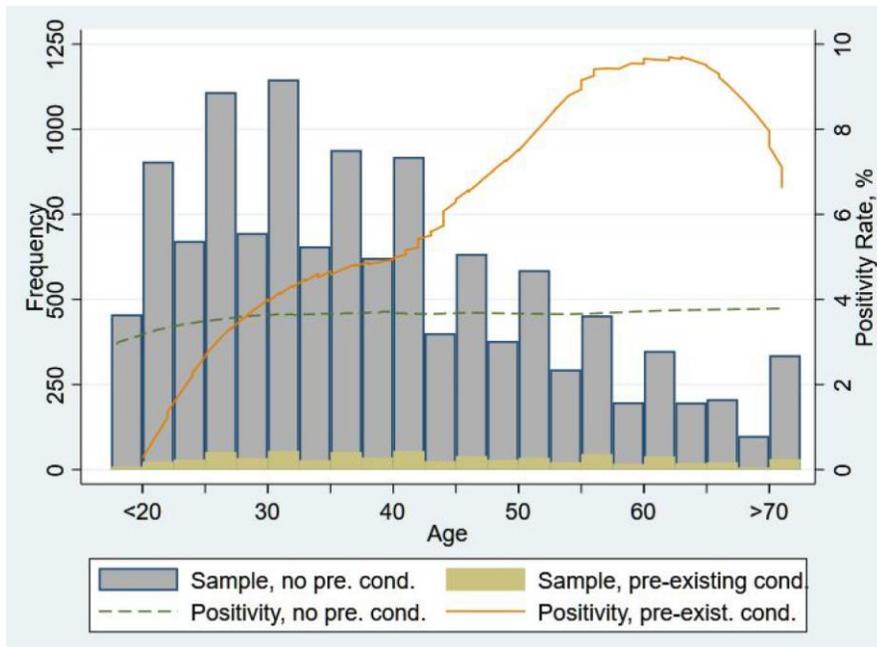


Figure 10: Age Profile and Positivity Rate by Pre-existing Conditions

4.5: What Predicts Covid-19 prevalence? A Multivariate Regression Analysis

While there is limited data available, we can still shed some light on what factors may be predictive of Covid-19 prevalence rates. In order to do so, we conduct a multivariate regression analysis utilizing some of the individual attributes mentioned above, as well as some group attributes (like the level of infection in your vicinity) and the population attributes (such as density) in your neighbourhood.

Table 2 below shows the result of this analysis. Column (1) first only includes variables that are not directly related to Covid-19 (i.e. whether someone is symptomatic and the positivity rate in a person's vicinity) since that is a prediction model that can be used without any testing or special surveying for Covid-19. Column (2) then adds additional factors that are related to Covid-19 specifically - namely whether an individual reports symptoms and the percentage of infected individuals nearby.

Column (1) shows that once we include all these factors together, the two factors that increase an individual's chances of being Covid-19 positive are (i) whether they have pre-existing conditions and (iii) the population density in their surrounding areas. Having a pre-existing condition increases the probability that one is Covid-19 positive by 2.2 percentage points (a 58.5% increase over the average positivity rate of 3.76% in our sample) while an increase of 100,000 people per sq-km (around 0.4 standard deviations increase in population density in our sample) increases the likelihood of infection by 0.3 percentage points for every 100,000 greater density.

Column (2) includes the additional factors that capture whether someone is symptomatic and the percentage of people who are infected in their vicinity. We find that a person is more likely to be infected if individuals near them are infected. People in a person's testing "pool" are those individuals who are essentially their neighbours or live close enough that they were part of the same pool of tests. Here we find a 1 percentage point increase in positivity rate in people who were in a person's testing pool increases the probability they are positive by 0.6 percentage points. Another way of saying this is if everyone in an individual's testing pool was positive then their probability of being Covid-19 positive would increase by 60 percentage points! Interestingly, there is an (additional) effect of 0.2 percentage points increase in being Covid-19 positive if there is a 1 percentage point increase in the percentage of people who are Covid-19 positive and live within a 1km radius of the (but not in their pool). However, there is no additional effect of the percentage of people that are between 1km to 2km, and 2km to 3km of where the person lives. These results suggest that the likelihood of infection increases the greater the percentage of people around an individual are infected but that this effect decays as the distance increases so that people more than 2 kms away don't matter as much, i.e., the infectivity likelihood is very localized. Interestingly, once we include these measures of local infection rates population density no longer matters. This suggests that higher density places are worse off only to the extent that a denser place means a greater chance that someone is infected nearby.

Regression analysis

VARIABLES	(1) Probability of Individual testing Positive	(2) Probability of Individual testing Positive
Age	9.56e-05 (0.000124)	5.67e-05 (0.000118)
Male	-0.00596 (0.00520)	-0.00373 (0.00495)
Symptomatic		0.00570 (0.00965)
Any Preexisting Condition	0.0216*** (0.00757)	0.0164** (0.00748)
Traveled Domestically	0.0109 (0.00888)	0.00550 (0.00850)
Traveled Internationally	-0.0381 (0.0673)	-0.0152 (0.0638)
Average Household Size	-0.000138 (0.00113)	-0.000144 (0.00108)
Population Density (x00,000 per sq km)	0.00306*** (0.000689)	0.000341 (0.000687)
Positivity Rate in Pool, %		0.00594*** (0.000184)
Positivity Rate (1km radius), %		0.00229*** (0.000647)
Positivity Rate (2km radius), %		-0.000697 (0.00153)
Positivity Rate (3km radius), %		0.00113 (0.00155)
Constant	0.0302*** (0.00986)	0.00522 (0.00955)
Observations	11,873	11,815
R-squared	0.003	0.106

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Note: Positivity rates in the pool and 1km, 2km, 3km radii are calculated using 'leave-one-out' method by not including the individual in the calculation. This is done to avoid spurious results where one's own positivity also contributes to the calculated positivity rate.

Table 2: What Predicts Covid-19 prevalence? A Multivariate Regression Analysis

4.6: Testing Costs and Efficiency

One of the concerns regarding testing may be the financial cost associated with it. While these costs may well be worth it given the value testing and prevalence detection has in reducing future spread, part of the value-additions of our smart sampling and testing strategy has been a large reduction in these costs and an increase in the efficiency of testing. Based on data received from P&SHD the typical cost of testing is Rs 3,500 per test. Once we factor in the personnel costs of actually collecting test samples this goes up to Rs 3,663 per test. In our case, we were able to adopt a pooled testing approach by creating pools of ten individuals. Therefore, while each individual person in the pool would have a sample taken, there would only be one test run for the combined sample. If the test was negative no further testing would be done. However, if the test was positive, then each of the ten individuals in that pool would be tested. Given that the majority of pools were negative (79%), this meant that far fewer overall tests needed to be conducted. As a result of this we were able to effectively reduce the cost to Rs 1,170 per test, or Rs 1,334 per test once we factor in costs of collecting the samples.

This is a remarkable saving, as pooled testing effectively reduces the overall cost of testing by a factor of three (i.e. a 63.5% to 66.5% drop in costs). Another way to think about this is that just this change alone allows us to sample three times as many individuals in the same budget. While this is already a significant saving, these costs could be further reduced if larger samples were pooled.

The testing strategy that was adopted also allows us to make considerable headway on two key fronts, which are not only important for Pakistan but also for understanding transmission globally. First, in countries like Pakistan, the number of Covid-19 cases is rising rapidly with increasing evidence of community transmission. Once community transmission is underway, the epidemic is harder to contain through contact tracing alone; some level of community surveillance is required. Second, there is growing evidence, consistent with our data that a large fraction of people infected with Covid-19 are asymptomatic, and although there is a dearth of studies from Pakistan, it appears that many of these patients are truly asymptomatic, rather than pre-symptomatic. The question we are then in a unique position to ask is: How does community transmission and the presence of asymptomatic individuals affect the community surveillance strategy for Pakistan?

We are concerned here both about effective epidemic control and the costs of testing. For effective epidemic control, identifying individuals early in their infectious stage can help break transmission chains. But if many infective patients are asymptomatic, screening on symptoms alone is insufficient. It can also be very costly. To see why, consider how much it costs to detect a single positive patient. With community surveillance, this includes both the costs of screening and the cost of testing. If a large fraction of patients is asymptomatic, screening on symptoms may be less effective even compared to simple random sampling.

In our testing data, we can calibrate these costs exactly. In doing so, we find that, even without pooling, a strategy of smart testing like the one we used is more than 50% cheaper in the cost

per positive case detected than a policy of first screening for symptomatic individuals and then testing (only) them. Once we factor in pooled testing, these savings are as high as 75%. Taken together, these results highlight how large a saving can be generated by adopting a process of smart sampling and testing.

5. Policy Recommendations

This study has shed light on a data-driven policy response to COVID-19. Until a vaccine is successfully produced and deployed—a process which may take several months even under optimistic scenarios—lowering COVID-19 prevalence should remain a central policy focus. Some policy recommendations, based on the findings of this study, are discussed in this section.

Health policies need to be revised to safeguard the elderly and women, who are at higher risk of infections. Given that the elderly have higher infection rates and worse morbidity outcomes, health policy needs to especially protect them. Moreover, given that women have higher infection rates than men, there needs to be a move towards greater testing for women.

Policy responses may be more effective if they are more localized. The substantial and very localized variation observed in Covid-19 prevalence rates suggests that lockdown policies may be more effective and less costly to the economy if they can be implemented in the smallest possible geographical and administrative units.

Community messaging campaigns should focus on universal prevention. Given the high degree of asymptomatic cases, safety messages need to stress universal prevention (wearing masks and washing hands frequently etc.) for all *regardless* of whether they have symptoms or not.

Frontline workers should be tested regularly. While healthcare frontline workers should be prioritized, there are a range of others who are also frontline workers - such as those working in pharmacies, exposed bureaucratic staff, emergency response teams, enforcement officials, supply chain logistics providers, immigration/border officials, public transport workers, retailers - and need to be tested.

Conduct follow-up surveys with all those who are detected positive. Healthcare needs and long-term morbidity outcomes should be traced out for the population.

An appropriately designed and systematically implemented contact tracing process can be critical in preventing spread, especially when there is a low level of infection. Setting up such a system will also help contain spread in places where there is risk of disease resurgence as has been seen during reopening of cities around the globe.

Continuation and expansion of the Smart Testing approach. Rigorous testing data as well as expanded data gathering exercises (including other health and socioeconomic outcomes) will

allow policymakers to understand prevalence and transmission of Covid-19, design effective policies to mitigate spread, and assess the tradeoffs involved in lockdown and reopening so they can mitigate the losses and help rebuild the economy and society.

6. Conclusion

The Smart Sampling and Testing process implemented in Punjab built an impressive and significant capacity in the P&SHD team to run smart and cost-effective sampling testing and data collection activities that remain as critical to informing the Covid-19 policy response as ever. The analysis above also revealed several useful and important findings that can directly inform the Covid-19 response for the months to come.

First, this study has demonstrated that the smart sampling and testing methodology, combined with pooled testing, is not only feasible but can be carried out regularly and at-scale across the province to determine how Covid-19 prevalence is changing over time, even at the level of sub-neighbourhoods within cities. This is a critical tool in helping us understand the effectiveness of various kinds of lockdown policies, as well as how soon and costly an opening up can be. Our analysis suggests that the initial lockdown was likely quite effective and such policies may continue to be an effective tool for containment. Importantly, by tying the extent of future lockdowns to such data, governments can ensure effective implementation of an evidence-based smart lockdown and containment policy.

Second, our results have shown that we can readily combine the testing process with a rapid and real-time data collection exercise. While we used the system to focus on data that helped us better understand the demographics and patterns of the disease, we can easily amend this and add data on other non Covid-19 related health outcomes as well as economic impact outcomes (such as food security, poverty, employment, income, etc.) that will help provide a sense of how costly the lockdowns are to the economic welfare of citizens. By obtaining this additional data, we will be able to arrive at a better policy decision that balances the costs and benefits of each lockdown/opening up strategy. These data will remain just as important even as Covid-19 prevalence diminishes since they would then help inform and better target the process of rebuilding and restoring affected families and the overall economy.

Third, our study has yielded important policy implications described in the section above. These policy recommendations show how a data-driven and evidence-based approach to designing Covid-19 response measures can yield a more appropriate and effective policy response.

Appendix A: SCALE Operational Plan³

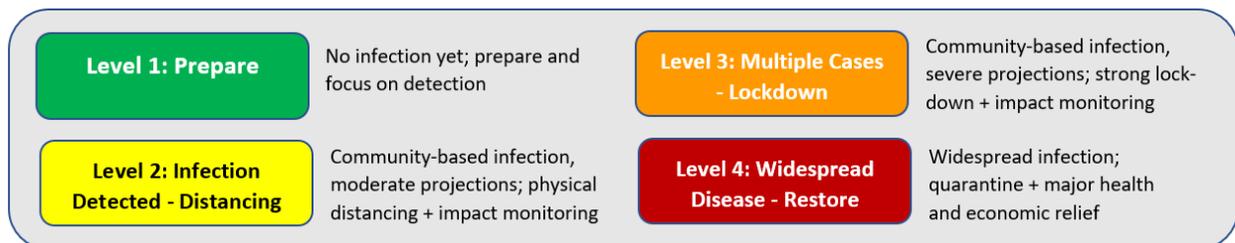
Governments around the world face two incredibly hard choices in responding to the COVID-19 pandemic: **Lockdown** and prevent spread, but risk economic collapse and potentially many dying from non-COVID reasons – *OR* – **start opening up** to minimize the socio-economic fallout, but risk COVID-19 resurgence. We propose a plan that helps make these choices in the most informed, feasible and timely way.

The problem we face is there is little data on potential health & economic impacts to figure out tradeoffs between choices and to design a real-time, effective, & evidence-based response

We propose an Active Learning & Refined Response Approach

The key principles underlying our approach are: 🌀 Data is Critical 🌀 Policy Responses Must Be Data-Driven & Data-Responsive 🌀 Policy Responses Must Vary with Local Conditions 🌀 Community Messaging is Key to Ensuring Voluntary Compliance 🌀 Response must be both Immediate and Sustained 🌀 Partner with on the Ground Implementers to Leverage Existing Capacity

Based on these principles, we have developed a standardized COVID-19 alert system and operational plan which allows for a staged time-varying response with rapid learning and refinement



³ Contributing team members include (in alphabetical order by last name): Dr Tahir Andrabi (LUMS / Pomona), Dr Matt Andrews (Harvard), Dr Ali Cheema (LUMS), Dr Jishnu Das (Georgetown / Center for Policy Research), Dr Adnan Q. Khan (LSE), Dr Mishal S. Khan (LSHTM), Dr Asim I. Khwaja (Harvard), Dr Farhan Majid (University of Georgia), Dr Aryn A. Malik (Yale), Anum Malkani (CERP), Dr Tyler McCormick (University of Washington), Dr Saad B. Omer (Yale), and Maroof A. Syed (CERP).

This operational plan lays out a district evidence-responsive response strategy that seeks to minimize disease spread and limit adverse socio-economic impact by targeting response to the current & projected level of prevalence in the smallest feasible area (S) within the District (D). Each set of actions is associated with a set of agents.

Stages	<p style="text-align: center;">Level 1 – PREPARE</p> <p style="text-align: center;"><i>No infection identified in <u>District (D)</u></i></p>	<p style="text-align: center;">Level 2 – INFECTION DETECTED in S</p> <ul style="list-style-type: none"> • <i>First case identified in <u>area (S) in the District (D)</u></i> • <i>Population has low impact and/or low health risk <u>vectors</u></i> 	<p style="text-align: center;">Level 3 – MULTIPLE CASES in S</p> <ul style="list-style-type: none"> • <i>Multiple cases identified in <u>area (S) in the District (D)</u></i> • <i>Population has high impact and/or high health risk <u>vectors</u></i> 	<p style="text-align: center;">Level 4 – WIDESPREAD DISEASE in S</p> <p style="text-align: center;"><i>Widespread disease in <u>area (S) in the District (D)</u></i></p>
Smart Testing & Data	<ul style="list-style-type: none"> • Sentinel surveillance • Syndromic surveillance to screen & test symptomatic patients and people with high impact & health risk vectors • Establish baseline data on standardized set of key health (service) indicators & mortality 	<p>Level 1 plus:</p> <ul style="list-style-type: none"> • Contact tracing & testing (required data) • Testing of front-line workers and high impact individuals • Syndromic Surveillance Robocalls to focus on infected S's • Follow-up Phone surveys of infected • Social mobility data to monitor compliance • Monitor changes in health indicators 	<p>Level 2 plus:</p> <ul style="list-style-type: none"> • Expand contact tracing • Start syndromic surveillance robocalls in adjacent S's • Survey of recovered 	<ul style="list-style-type: none"> • Test primarily for therapeutic purposes in infected S • Surveillance using seroprevalence surveys • Collect additional data on impact of morbidity and health sector constraints to meet Level 4 needs • Continue monitoring changes in health indicators, broaden set of health indicators
Physical Distancing Measures	<ul style="list-style-type: none"> • Standard health practices (washing, masks) • Protective Organizational & Management practices • Light physical distancing measures • District distancing 	<p>Enhance physical distancing by additional measures (on top of level 1):</p> <ul style="list-style-type: none"> • S-unit distancing • Isolate confirmed cases, move severe to ICU • Quarantine contacts • Enact enforcement & protection 	<p>Implement stronger measures & lockdown with additional measures (on top of level 2):</p> <ul style="list-style-type: none"> • Move infected S's into full lockdown • Declare adjacent S's to be in Level 2 • District Quarantine 	<ul style="list-style-type: none"> • Consider easing lockdown in S based on seroprevalence survey data (continue physical distancing & self-quarantine for high-risk people) • Declare Level 3 in adjacent S's
Community Messaging	<ul style="list-style-type: none"> • Consistent with Physical Distancing measures and Community Messaging principles 	<p>Level 1 plus:</p> <ul style="list-style-type: none"> • Public Information on Prevalence • Public Information on Support • Based on data, communicate messages to support appropriate healthcare seeking behavior 	<ul style="list-style-type: none"> • Same as Level 2 and expand to entire D 	<ul style="list-style-type: none"> • Same as Level 3, but emphasize severity • Emphasize health & relief response • Special expedited helpline for Level 4 S's
Actionable Decisions	<ul style="list-style-type: none"> • Regularly monitor the situation & determine alert level • Prepare contingency plans • Prepare for changes in healthcare service delivery 	<p>Monitor the situation and refine response:</p> <ul style="list-style-type: none"> • Based on social mobility data, increase messaging and enforcement if needed • Use testing data to monitor infection rate & determine alert level • Use data to assess adverse impacts, including wider health impacts, & target health & socio-economic support • Adapt healthcare service delivery as needed 	<ul style="list-style-type: none"> • Same as Level 2, but expand to entire D • Prepare for expanded medical, food and other needs • Draw on dedicated external and internal support - e.g. army, volunteers 	<p>Level 3 plus:</p> <ul style="list-style-type: none"> • Scale up support resources and welfare measures to address morbidity impact • Ensure Level 4 S's needs are being heard and met

Appendix B: SCALE Smart Testing Strategy⁴

Step 1: Define the lowest sampling grids “S” for City/Districts and conduct data diagnosis

- ‘S’ should be as small as minimally feasible (200+ households)
- Shape files (geo-referenced boundaries) will be required for ‘S’
- Ideal ‘S’ are census block/mauzas as that will allow for overlaying census information (population density, distribution of age etc.)
- Alternative options: If census blocks/mauzas not available, then the following can also be used:
 - ◆ UCs
 - ◆ Geo-referenced grids
 - ◆ Creating geo-referenced polygons using cell phone towers
- Conduct a data diagnosis to assess the usability of the data⁵ that is being collected and understand its workflow

Inputs	Outputs
<ol style="list-style-type: none">1. Shape files of ‘S’ grids2. Census data on ‘S’ grids3. Geo-locations of isolation centers, quarantine facilities hospitals and labs conducting PCR Tests4. Data Diagnosis	N/A

Step 2a: Contact Tracing Testing

- Identifying contacts of known positive cases - there are two ways to obtain this:
 - ◆ First preference: Through Call Detail Records (CDR) and/or Cell Phone Tower Data. The steps involved would be as follows:

⁴ The lead contributors for this document include Dr Ali Cheema (LUMS), Dr Jishnu Das (Georgetown / Centre for Policy Research), Dr Adnan Q. Khan (LSE), Dr Asim I. Khwaja (Harvard), Dr Farhan Majid (University of Georgia), Dr Ameen A. Malik (Yale), Dr Tyler McCormick (University of Washington), Dr Saad B. Omer (Yale), Omer Qasim (CERP), Maroof Syed (CERP).

⁵ Health, location, and contact tracing data is personal data. Information collectors should ensure that the data collected is accurate and secure. The integrity of data can be improved by cross-referencing it with reputable databases and by providing access for the consumer to verify it. Information collectors can keep data secure by protecting against both internal and external security threats. They can limit access within their organization to only necessary employees to protect against internal threats, and they can use encryption and other computer-based security systems to stop outside threats.

- Provide telecom operators with cell phone numbers of infected patients
 - Identify exposed people - identify individuals that were in the same physical location (same tower) during the span of 15 minutes and/or 30 minutes overlapping window in the past 14 days (ideally 3 weeks). If this number is not too large, use it. If it is too large, narrow it down further to exposed *contacts*
 - Define contact list - ask telecom operators to identify all numbers called by the infected person's number for at least 14 days (ideally 3 weeks) prior to the person's infection start date
 - Narrow down list of exposed *people* to exposed *contacts* by excluding everyone from exposed people list who are not in contact list
- ◆ Second preference: Actual contact tracing done through in-person surveys/phone calls/robocalls (will capture contact details and work/home location of contacts)
- Test & administer basic in-field survey (survey captures basic demographics useful in refining disease model) to contacts and geotag where they live/work and assign to appropriate 'S' grid

Inputs	Outputs
<ol style="list-style-type: none"> 1. Individual level data on contact tracing (CDR or captured through surveys/robocalls) 2. Home and work geo-locations of traced contacts 	<ol style="list-style-type: none"> 1. Overlaying the data from inputs on appropriate 'S' grids

Step 2b: Exposed Frontline Worker Testing

- Identify all exposed frontline workers. These are workers (medical staff, caretakers, police and government officials, retailers, etc.) who are/have been in direct contact with infected people (with 10 minutes or more of exposure)
- Test & administer in-field for contacts and geotag where they (i) work and (ii) live and assign to appropriate 'S' grids

Inputs	Outputs
<ol style="list-style-type: none"> 1. Individual level data on exposed front-line workers 2. Home and work geo-location of frontline workers 	<ol style="list-style-type: none"> 1. Overlaying the data from inputs on appropriate 'S' grids

Step 3: Testing in 'S' grids

- Prioritize PCR testing in 'S' grids of the following types - start testing in Category 1 first and then move to Categories 2 and 3:
 - ◆ Category 1
 - 'S' grids that have no infection detected yet but are next to a grid which has an infection
 - 'S' grids that have a high number of frontline workers working or living in them
 - 'S' grids with people who have high mobility/travel/connectivity
 - ◆ Category 2:
 - 'S' grids that have high population density
 - 'S' grids with large fraction of high health risk people (elderly etc.)
 - ◆ Category 3:
 - 3a: all remaining uninfected grids
 - 3b: all grids that already have been detected with high infection (PCR Testing in such areas is not as informative but these grids are top priority for seroprevalence testing)

- Conduct the following types of prevalence testing:
 - ◆ **Screening and Testing:**
 - Ideal - but may be in limited supply; Instead of randomly testing people in 'S' grid, test those who are High Infection Susceptibility (i.e. people who have higher likely of infection) since this gives a more efficient way of testing (i.e. will need to use less tests to detect infection). These can be defined as people with (i) high number of physical contacts/interactions and/or (ii) high mobility
 - For PCR Testing, pool 10 or 20 tests each depending on what you have been instructed to do. Follow Pooled Testing protocol. Make sure everyone is also administered an in-field basic survey
 - **Preferable method is using CDR data:**
 - High physical contacts: The following complementary approaches can be used
 - ◆ run algorithm for each phone number by seeing in one week (can use the most recent full week - make sure weekend and weekdays are included) what is the total number of OTHER phone numbers that were in the same physical location as they were in an overlapping 30 minute time window - list (phone no & location) of the top 5% percentile in this measure
 - ◆ run algorithms on CDR data to identify high contact nodes who are in contact with more people/links (restrict these to

those within the same city and also see which of their edges are more active for further contact tracing)

- High mobility folk: These are people who move around a lot (so could be more likely to get exposed) - run algorithm to count how many unique towers (i.e. different locations) the person has been at - can refine this over time to maybe weigh locations that are hot spots - provide list (phone & location) of the top 5% percentile in this measure
- Build profiles of high contact/mobility individuals by combining CDR data with administrative records
- **If CDR data not available, then define High physical contacts/mobility by:**
 - Randomly arriving at the GPS pin location provided in 'S' Grid - Test & Survey member of household who has highest contact/travel in past 14 days (may not be owner but could be domestic help who does regular shopping etc.) AND ask the household head who is a person with high contacts nearby OR who is their friend nearby (people who are identified as friends will likely have higher connectivity) and then go and Test & Survey them
- ◆ **Sentinel Surveillance PCR Testing**
 - Identify main 2-3 sewage collection points in 'S' Grid and collect sample by using protocols similar to those for polio environmental surveillance through sewage sampling
 - This is an experimental approach and will require a pilot before being rolled out on a broader scale. Areas where grid maps of sewerage are available would be required
 - Take samples and note the GPS location where the sample was taken
- ◆ **Phone-based Syndromic Surveys**
 - Call randomly selected numbers (if possible, prioritize numbers that show high mobility or physical contact in past 2-3 weeks) and administer survey of symptoms on the phone (this can also be done through robo-calls)

Inputs	Outputs
<ol style="list-style-type: none"> 1. Categorize 'S' Grids <ol style="list-style-type: none"> a. Identify high mobility and high contact individuals to administer appropriate tests 2. Contact details and biographic data on high mobility/high contact individuals 	<ol style="list-style-type: none"> 1. COVID19 detection PCR tests and field survey on people with high mobility and high no of contacts 2. Sentinel Surveillance by testing water samples and recording GPS location where the sample was collected from 3. Syndromic Surveys through Robocalls and Phone CATI

	<ol style="list-style-type: none"> 4. Positive cases overlaid on respective 'S' grids 5. Action items for stakeholders to take in respective 'S' grids <ol style="list-style-type: none"> a. SOPs for dealing with Positive cases (option for isolation at home)
--	--

Step 4: Follow-up socio-economic and broader health surveys

- All individuals tested are administered short phone surveys that will capture their socio-economic as well as broader health situation. This will be used to see what other costs they are incurring and what help may be useful to provide to them

Inputs	Outputs
<ol style="list-style-type: none"> 1. Contact details and relevant biographic data of all individuals tested 	<ol style="list-style-type: none"> 1. CATI phone surveys to capture socio-economic and broader health concerns 2. Overlay the information on appropriate 'S' grids 3. Devising action items for the stakeholders in 'S' grids

Appendix C: Testing and Tracing Survey Instrument⁶

This repository consists of 7 modules.

I. TESTING SURVEY: Module 1, 2, 3, & 4 will be asked from the same respondent at the time of taking the sample for PCR test. This will be done for 10 respondents on each GPS point for a particular area (S Grid number).

Module 1 - *S Grid Cluster Mapping / Identification*: This section is to keep track of areas covered (S grid and GPS points).

Module 2 - *Respondent Identification and Testing* : This section contains basic respondent information and where test sample should be taken.

Module 3 - *Travel History* : This section records basic travel history of the respondent.

Module 4 - *Clinical History* : This section records clinical history of the respondent.

II. FOLLOW-UP SURVEYS: The modules below are 3 different follow-up survey instruments for Pool Testing follow-up (in case the result is positive), a follow-up with individual positive cases & a Contact Tracing questionnaire.

Module 5 - *Pool Testing Follow-Up* : If for any specific pool, the result comes out positive, then follow-up individual test samples will be taken for each individual that was part of that pool.

Module 6 - *Confirmed Positive Case(s) Follow-Up* : Here, enumerators will survey people from the pool who tested positive. They will gather information on other people these positive cases have come into contact with.

Module 7 - *Contact Tracing* : Enumerators will reach out to the contacts provided in Module 6 and survey these contacts using questions here. To get additional data points on traced contacts, Module 3 (Travel History) and Module 4 (Clinical History) are also recommended to be administered.

⁶ The lead contributors for this document include Dr Ali Cheema (LUMS), Dr Jishnu Das (Georgetown / Centre for Policy Research), Dr Adnan Q Khan (LSE), Dr Asim I Khwaja (Harvard), Dr Farhan Majid (University of Georgia), Dr Aymn A. Malik (Yale), Dr Tyler McCormick (University of Washington), Dr Saad B. Omer (Yale), Omer Qasim (CERP), Maroof Syed (CERP).

SCALE Testing & Tracing Survey Instrument

Module 1 - S Grid Cluster Mapping / Identification

Input Field Name	Question	Relevance	Option Choices	Notes
S 1	S Grid number (Census block number)			<i>This will be manually entered by referring to the S grid numbers</i>
S 2	Choose one of the following to describe this survey		1. First interview at pin location 2. Tracked contact at PIN location 3. Follow up after someone tested positive	
S 3	Which area is this?	If S 2 = 1		
S 4	Which GPS/ PIN point is this?	If S 2 = 1		<i>This will be manually entered by referring to the GPS point numbers within each S grid</i>
S 5	How many people have you surveyed at this PIN location, including this survey?	If S 2 = 2		

SCALE Testing & Tracing Survey Instrument

Module 2 - Respondent Identification and Testing

Input Field Name	Question	Relevance	Option Choices	Notes
RI 0	Identify the person in this Household who has the highest number of contact/physical interaction with other and/or is the most mobile?			
Note for enumerator: Record the following for the identified person				
ID	ID (Auto-generated)			<i>This is a unique in- house generated ID that will be unique for each individual tested/surveyed</i>
RI 1	First Name			
RI 2	Last Name			
RI 3	Gender			
RI 4	DOB			
RI 5	CNIC Type (Self/Guardian)		1. Self 2. Guardian	
RI 6	CNIC			<i>To be entered without dashes or spaces</i>

RI 7	Contact number (Mobile)			<i>To be entered without dashes or spaces</i>
RI 8	Alternate number			<i>To be entered without dashes or spaces</i>
RI 9	District/City			
RI 10	Home Address			
RI 11	Home geolocation			
RI 12	Work Address			
RI 13	Work geolocation			
RI 14	Occupation		1. Health Occupation 2. Self- employed (owns his/her own business) 3. Law Enforcement and armed forces (police, army, etc.) 4. Legal Profession 5. Employed 6. Teacher/ Professor 7. Student 8. Unemployed 9. Other	
RI 15	Name of company/org			
R16	During an average day in the last two weeks, how many people have you spoken to for at least 10 minutes?			<i>Write a number here</i>
R17	Over the last two weeks, have you been to any place with more than 10 people (this could include a market, a wedding, or a masjid).		1. Yes 2. No	
Note for enumerator: After recording the above information, conduct the test for pool sample				
PL 01	Pool or Individual Sample		1. Pool 2. Individual	
PL 02	Pooling ID (Auto-generated)	if PL01=1		<i>Unique ID for pooled test sample generated at the backend</i>
PL 03	Test ID (Auto-generated)	if PL01=2		<i>Unique ID for individual test sample generated at the backend</i>

PL 04	Date of sample			<i>Record date</i>
Note: After taking sample, ask the following:				
FU 01	Name of your close friend in this neighborhood?			
FU 02	Name of person1 in this neighborhood with highest contact points/mobility?			
FU 03	Name of person2 in this neighborhood with highest contact points/mobility?			
FU 04	Name of person3 in this neighborhood with highest contact points/mobility?			

SCALE Testing & Tracing Survey Instrument

Module 3 - Travel History

Input Field Name	Question	Relevance	Option Choices	Notes
ID	ID (Auto-generated)	Pull from central database		<i>This is a unique in-house generated ID that will be unique for each individual tested/surveyed</i>
TH 01	Travel History in the last 3 weeks		1.Domestic 2.International 3. No travel	<i>Select multiple</i>
TH 02	Which country did you travel to?		1. United Kingdom (UK) 2. United States of America (USA) 3. China 4. United Arab Emirates (UAE) 5. Thailand 6. Turkey 7. Germany 8. Qatar 9. Oman 10. Bahrain 11. Saudi Arabia 12. Other	
TH 03	Did you self-isolate for 14 days when you returned to Pakistan?	PI 6 = 2	1. Yes 2. No	

TH 03	Point of entry into Pakistan	PI 6 = 2	1. Lahore 2. Islamabad 3. Karachi 4. Sialkot 5. Multan 6. Peshawar 7. Quetta 8. Other	
TH 04	Which areas did you visit during domestic travel? (select all that apply)	PI 6 = 1	1. Islamabad 2. Karachi 3. Quetta 4. Peshawar 5. Sukkur 6. Bahawalpur 7. Lahore 8. Faisalabad 9. Gujranwala 10. Gilgit 11. Other	

SCALE Testing & Tracing Survey Instrument

Module 4 - Clinical History

Input Field Name	Question	Relevance	Option Choices	Notes
ID	ID (Auto-generated)	Pull from central database		<i>This is a unique in-house generated ID that will be unique for each individual tested/surveyed</i>
CH pec	Do you have any pre-existing conditions?		1. Diabetes 2. Hyper-tension 3. Obesity 4. Cancer 5. Smoking 6. Cardiovascular disease 7. Chronic lung disease 8. Chronic liver disease 9. Chronic renal disease 10. Malignancy 11. Other	

CH 1	Symptoms		1. Chills 2. Vomiting 3. Nausea 3. Diarrhea 4. Headache 5. Rash 6. Conjunctivitis 7. Muscle Ache 8. Joint Ache 9. Loss of appetite 10. Nose bleed 11. Fatigue 12. Seizures 13. Altered Consciousness 14. Loss of smell 15. Loss of taste 16. Fever 17. Other neurological signs 18. Other symptoms	
CH 2_num	How many?			
CH 3_num	How many confirmed cases have you been in contact with?			

SCALE Testing & Tracing Survey Instrument

Module 5 - Pool Testing Follow-Up (Only to be done for a positive pool)

Input Field Name	Question	Relevance	Option Choices	Notes
PL 02	Pooling ID (Auto-generated)	Pull from database if PL01=1		
ID	ID (Auto-generated)	Pull from central database		<i>This is a unique in-house generated ID that will be unique for each individual tested/surveyed</i>

Note: After pulling this data, take the sample for testing

Test ID	Test ID			<i>Unique ID generated at the backend (if, in section 2 'Respondent info and testing' Item PL1 was 'Pool') Otherwise if PL1 was 'individual', then pull data)</i>
---------	---------	--	--	---

SCALE Testing & Tracing Survey Instrument

Module 6 - Confirmed Positive Case(s) Follow-Up

Input Field Name	Question	Relevance	Option Choices	Notes
Sample ID	ID (Auto-generated)	Pull from central database, assigned during pool testing follow up (Stage 5)		<i>This is a unique in-house generated ID that will be unique for each individual tested/surveyed</i>
CI 1	How many people have you been in contact with?			<i>Provide hint for enumerator from clinical history questions CH 2_num and Ch 3_num. This is to make sure that the respondent does not understate his number of contacts once he/she knows they are positive</i>
Repeat Group (count = CI 1)				
CI 1_1	First Name of person			
CI 1_1_1	Last Name of person			
CI 1_2	Mobile phone number			
CI 1_3	PTCL phone number			
CI 1_4	City/ Area of infection			
CI 1_5	Work GPS location/ address			
CI 1_6	Home GPS location/ address			
CI 1_7	What was the nature of your contact with this person?		<ol style="list-style-type: none"> 1. Shook hands with him/her 2. Hugged him/her 3. Prolonged interaction (>10 mins) with him/her in an enclosed space (room, mosque, office, etc) 4. Prolonged interaction (>10 mins) in an open space (market, park, etc) 5. Brief Interaction (<10 mins) with him/her in an enclosed space 	
			6. Brief interaction	

			(<10 mins) with him/her in an open space	
End Group				
SCALE Testing & Tracing Survey Instrument				
Module 7 - Contact Tracing (from contacts of confirmed positive cases)				
Input Field Name	Question	Relevance	Option Choices	Notes
Sample ID	Sample ID	Pull from central database, assigned during pool testing follow up (Stage 5)		<i>This is a unique in-house generated ID that will be unique for each individual tested/surveyed</i>
Contact Information				
RI 1	First Name			<i>You will have some of this information such as first name, last name from the "Follow-up with positive cases." This can be pulled whereas the rest will be recorded. The idea here is to repeat the exercise with these contacts and maintain a central database of all respondents.</i>
RI 2	Last Name			
RI 3	Gender			
RI 4	DOB			
RI 5	CNIC Type (Self/Guardian)			
RI 6	CNIC			
RI 7	Contact number (Mobile)			
RI 8	Alternate number			
RI 9	District/City			
RI 10	Home Address			
RI 11	Home geolocation			
RI 12	Work Address			
RI 13	Work geolocation			
RI 14	Occupation		1. Health Occupation 2. Self-employed (owns his/her own business) 3. Law Enforcement and armed forces (police, army, etc.) 4. Legal Profession 5. Employed 6. Teacher/ Professor 7. Student 8. Unemployed 9. Other	
RI 15	Name of company/ organisation			

CT 1	According to our data [person name] has tested positive for COVID - 19, we know you may have come into contact with the above mentioned person. Please tell us what is your relation to above mentioned person. In what capacity do you know him?		<ol style="list-style-type: none"> 1. Immediate Family Member 2. Relative 3. Neighbour 4. Friend 5. Colleague 6. Other (specify) 	
CT 1_desc	If other, please specify			
CT 2	What was the nature of your contact with [person name] who has tested positive for COVID-19?		<ol style="list-style-type: none"> 1. Close contact (>10 mins) with [person name], who was not wearing face mask 2. Close contact (>10 mins) with [person name], who was wearing a face mask 3. Close contact (<10 mins) with [person name], who was not wearing a face mask 4. Close contact (<10 mins) with [person name], who was wearing a face mask 	
CT 4	How many people (other than the person they are being traced from) do you know who have tested positive for COVID- 19?	If CH 3 = No		
CT 5	How many people (who you personally know) do you suspect are positive (showing symptoms)?	If CH 3= yes		
CT 6	How many people have you come into contact with since you met [person the are being traced from]?			

Repeat Group (count = CT 6)				
CT rg 1	Contact Number			
CT rg 2	Alternate Contact Number			
CT rg 3	Home address			
CT rg 4	Work Address			
CT rg 5	Place of work (company, organisation etc.)			
CT rg 6	Geo Location (longitude)			
CT rg 7	Geo Location (latitude)			
End Repeat Group				
CT 7	What preventive measures have you adopted to mitigate the risk of contracting COVID-19? Select all that apply.		<ol style="list-style-type: none"> 1. Use of Face Mask 2. Handwashing with soap for 20sec 3. Use of alcohol based Hand sanitizer 4. Social Distancing 5. None of the above 	
CT 8	What are the preventive measures taken by your employer for the safety of the staff at the workplace? Select all that apply.		<ol style="list-style-type: none"> 1. Use of disinfectant for cleaning 2. Provided hand sanitizer 3. Raising awareness about preventative measures 4. Provided masks 5. Provided gloves 6. Allowed work from home 7. I am not going to the office/my office is closed 	
CT 9	Do you and your colleagues at the workplace follow the preventive measures of personal hygiene and social distancing?		<ol style="list-style-type: none"> 1. Do not follow at all 2. Do not mostly follow 3. Sometimes follow and sometimes don't follow 4. Mostly follow 5. Strongly follow 	

CT 10	Do you and your family follow the preventive measures of personal hygiene and social distancing?		<ol style="list-style-type: none"> 1. Do not follow at all 2. Do not mostly follow 3. Sometimes follow and sometimes don't follow 4. Mostly follow 5. Strongly follow 	
CT 11	In the last two weeks how many religious (friday prayers, congregation, etc) or social (marriage, party, etc) gatherings have you attended?		<ol style="list-style-type: none"> 1. None 2. 1 - 3 3. 3 - 10 4. More than 10 	
CT 12	In the last two weeks how many religious (friday prayers, congregation, etc) or social (marriage, party, etc) gatherings have you attended?		<ol style="list-style-type: none"> 1. Friday Prayers 2. Congregation prayers 3. Other religious gatherings (Naats, Milaad, Funeral, etc) 4. Conferences 5. Marriage Ceremony 6. Social Gathering (party, dinner, etc) 	
Repeat Roster (all option choices chosen in CT 12)				
CT 13	Area where you attended the event			
CT 14	City where you attended the event			
End Repeat Group				

Bibliography

Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M.D. 2020. Optimal targeted lockdowns in a multi-group SIR model. NBER Working Paper, 27102.

Alvarez, F.E., Argente, D. and Lippi, F. 2020. A simple planning problem for covid-19 lockdown. NBER Working Paper, 26981.

Anderson, R.M., Heesterbeek, H., Klinkenberg, D. and Hollingsworth, T.D. 2020. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet*, 395 (10228): 931-934.

Cherif A, Grobe N, Wang X, Kotanko P. 2020. Simulation of Pool Testing to Identify Patients With Coronavirus Disease 2019 Under Conditions of Limited Test Availability. *JAMA Network Open*. 2020;3(6):e2013075. doi:10.1001/jamanetworkopen.2020.13075

Eichenbaum, M.S., Rebelo, S. and Trabandt, M. 2020. The Macroeconomics of Testing and Quarantining. NBER Working Paper, 27104.

Hausmann, R., 2020. "Flattening the COVID-19 curve in developing countries", Project Syndicate. Available at: <https://www.project-syndicate.org/commentary/flattening-covid19-curve-in-developing-countries-by-ricardo-hausmann-2020-03>

Hausmann, R. and Schetter, U. 2020. Horrible Trade-offs in a Pandemic: Lockdowns, Transfers, Fiscal Space, and Compliance. CID Faculty Working Paper No. 382.

Keogh-Brown M. R., Smith R. D. 2008. The economic impact of SARS: how does the reality match the predictions? *Health Policy*, 88(1), pp.110-120.

Majid F, Omer SB, Khwaja AI. 2020. Optimising SARS-CoV-2 pooled testing for low-resource settings. *Lancet Microbe*. 2020;1(3):e101-e102. doi:10.1016/S2666-5247(20)30056-2

Manski, C., 2013. *Public Policy in an Uncertain World: Analysis and Decisions*, Harvard University Press.

Mobarak, A. and Barnett-Howell, Z. 2020. Poor Countries Need To Think Twice About Social Distancing. *Foreign Policy*. Available at: <https://foreignpolicy.com/2020/04/10/poor-countries-social-distancing-coronavirus>

Nour, M., Alhajri, M., Farag, E., Al-Romaihi, H., Al-Thani, M., Al-Marri, S. and Savoia, E. 2017. How Do the First Days Count? A Case Study of Qatar Experience in Emergency Risk Communication during the MERS-CoV Outbreak. *International Journal of Environmental Research and Public Health*, 14(12), p.1597.

Smith, R., Keogh-Brown, M., Barnett, T. and Tait, J. 2009. The economy-wide impact of pandemic influenza on the UK: a computable general equilibrium modelling experiment. *BMJ*, 339, pp.b4571.

Stock, J. 2020. Reopening The Coronavirus-Closed Economy. Working Paper no. 60. Hutchins Center on Fiscal and Monetary Policy at Brookings.

The International Growth Centre (IGC) aims to promote sustainable growth in developing countries by providing demand-led policy advice based on frontier research.

Find out more about our work on our website
www.theigc.org

For media or communications enquiries, please contact
mail@theigc.org

Subscribe to our newsletter and topic updates
www.theigc.org/newsletter-signup

Follow us on Twitter
[@the_igc](https://twitter.com/the_igc)

Contact us
International Growth Centre,
London School of Economic and Political Science, Houghton Street, London WC2A 2AE

IGC

**International
Growth Centre**

DIRECTED BY



FUNDED BY



Designed by soapbox.co.uk