CENTRE FOR
POLICY
RESEARCH

Photo Credit: Business Today

Testing for the Coronavirus at the State Level

NEELANJAN SIRCAR | JISHNU DAS
PARTHA MUKHOPADHYAY

21 April 2019

This brief discusses desirable strategies and important considerations to implement coronavirus (COVID-19) testing and data analysis. The suggestions are consistent with a modern statistical understanding of how to detect infection and understand where and how much it has spread across each state in India and, most importantly, minimise its transmission. All this has to be done in an environment of scarcity of tests. **India's level of testing** (in terms of tests per million) **is one-hundredth that of levels seen in countries such as Italy and Spain**.

Considerations

1. Testing of High-Risk Individuals

A good strategy would focus testing on those with high probabilities of having the infection, particularly given the limited number of tests. It will disproportionately test those who are most likely to have been infected. Of particular concern are people with high levels of interaction during lockdown/restricted periods that have the capacity to spread the virus quickly, **which includes health workers and police.** This is particularly a concern due to the asymptomatic transmission of COVID-19.

2. Detection of Infected Areas

A good strategy will look for cases where they are most likely to be found. Simply taking a representative sample of the population and testing is likely to find extremely few individuals. To see why, notice that even if India had 1 million infected persons (significantly more than the number detected in the United States, which has by far the most detected cases at the time of writing), it would still only be about 0.07% of the population. The chances of finding an infected person from random sampling is very low. Places more likely to see significant spread include those places that have already registered cases, but also those areas with higher population density, cramped housing, and poorer sanitation.
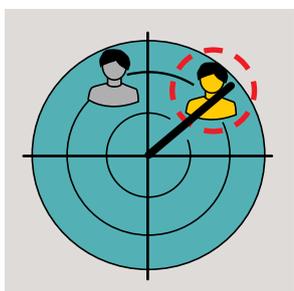
3. Early Detection Strategies

A good strategy finds infected persons in areas **where the virus has yet to spread significantly**. Employing a pure "hotspot" strategy is about responding after significant spread (e.g., the case of Tablighi Jamaat), while the intent is to anticipate where the virus might be. A small number of infected cases will become a "hotspot" a week later, so early containment is necessary. Responding after the fact only works in special cases. The much-touted "Bhilwara model" works because it does not have the same exposure in terms of migration and international travel as major cities in India – so spread is likely restricted to a small set of easily traceable clusters.

4. Estimation of Places of Hidden Infections

A good strategy permits for evaluation of risks of infection across the state, even in places not explicitly tested. This allows the government to understand the extent of infection across state and craft an optimal policy response to prevent things from getting out of control. The use of randomisation in the selection of where to look for cases allows for making such estimations with more accuracy. Furthermore, the collation of existing data, with high quality statistical analysis of data will allow for efficient processing of such information.

Designing a Testing Strategy

This section details a "hybrid strategy", with three key steps. It is consistent with ICMR's (Indian Council of Medical Research) testing protocols, as announced periodically:

1 Testing all individuals who may become "super-spreaders" – i.e., those who may plausibly be infected and have high levels of population interaction, even in lockdown/restricted periods;

2 Randomisation of areas to test, with a focus on the riskiest areas, and a combination of methods to find as many infected persons as possible in each area. This randomly samples areas to test based on risk-level and then locates infected persons based on a series of screening mechanisms and local reports, and;

3 Iteratively re-sampling after each round of tests based on inputs from real-time, high-quality analysis of incoming data. This means an earlier stage informs testing sample decisions in a later stage. The release of **anonymised** data in the public domain about all people who are tested will help crowd-source analysis about the possible determinants of infection probability.

This approach has the following benefits:

- It minimises the risk of transmission from super-spreaders.

- The use of randomisation in the selection of areas, with a focus on the riskiest areas, allows for analysis to detect hidden infections and to calibrate policy for future rounds of testing.

- The use of a medley of techniques to find infected individuals in a selected area maximises the detection of infected individuals.

- The iterative testing approach allows the government to save precious medical and testing resources in detecting coronavirus infections.

1. Detecting Key Spreaders:

The focus needs to be on identifying those with the greatest capacity to spread the virus. A major goal in testing needs to be the identification of infection among those who interact with many others. **In short, the goal of testing and containment needs to be more than identifying who *has* the virus – it needs to look towards who is most likely to *spread* the virus.**

We need tests on the most sensitive individuals to be processed quickly ("high throughput") and reliably. As such, we suggest conducting PCR (polymerase chain reaction) testing for the following individuals as quickly as possible:

(A) **All healthcare workers,** starting with those handling COVID-19 cases or cases of respiratory distress should be tested. They are the most exposed and therefore more likely to be infected. If the infection level in this group is low, the infection level in the population is likely to be very low (though since they are also trained to take precautions and have access – sometimes – to protective equipment, their chances of catching an infection may be reduced substantially). According to the ICMR protocol (dated 20 March), *"Asymptomatic direct and high-risk contacts of a confirmed case should be tested once between day 5 and day 14 of coming in his/her contact"*
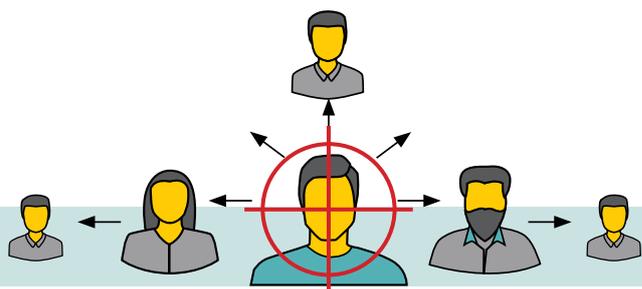
(b) **Essential services' workers,** who have been allowed to move during the lockdown, should be tested for the same reason as healthcare workers (possibly in the last week of the lockdown, to maximise the chances of finding infected persons). This is especially true of the police, given the methods of police control and the mixing between on-duty and off-duty personnel. If the infection level in this group is low (also, since they do not have the

training and equipment of healthcare staff, their chances of catching an infection is more similar to that of the general population), the infection level in the population is likely to be very low.

(c) **Immigration service staff and airline crew** at the airport are also very susceptible, but India has banned international travel for almost two weeks now – infections among such persons should have emerged, i.e., they should have become symptomatic and sought medical help. If so, they should be tested as per existing protocol. If operations resume, infected crew would present a major threat to passengers. Ideally, one should screen them for asymptomatic presentations.

However, under current ICMR guidelines (dated 4 April), asymptomatic persons cannot be tested, except in areas reporting clusters (containment zone) and in large migration gatherings/evacuees centres. So, for (b), (c) and for health workers who are not *direct and high-risk contacts of a confirmed case*, will be considered as a high risk individual and prioritised for testing.

2 . Selecting Areas to Test:

The strategy for testing involves disproportionately testing in the most risky areas, and using a combination of techniques to find infected persons in these areas. *It is worth reiterating that waiting for an area to become a big "hotspot" means that, in containment terms, it is already too late.* There are huge costs if one waits till there is significant spread. Yet, it is difficult as yet, with what is known, to guide our selection of areas. This has to be an iterative process, as described below.

The goal of any larger testing protocol needs to be efficient detection of cases across the state, especially emerging clusters – so they can be contained quickly without the heavy use of resources. The oft-discussed "Bhilwara model" is dependent on low rates of infection, highly clustered in one area with less interaction outside of the city. Simply relying on this strategy (and allowing "hotspots" to develop before containment) could impose very heavy costs on a state-wide level.

In order to select areas to test, we suggest including as much easily available information to engage in "targeted" testing of the population, to select the populations with the highest risk of infection, with the help of occupational markers (health workers, police, emergency services, flight crew, airline and airport staff, etc.). Already, we see the effects of infected healthcare workers in Mumbai.

Sampling Unit: The list of polling booths, and the voters contained therein, provides an appropriate basis upon which to conduct rapid sampling. The quality of detection hinges on the precision of geographic information and this will enable individuals to be geographically located at a fine granular level. Even though there may be many residents of various cities and villages not on the voter list, broadly, the most densely populated areas will have the most polling booths – that is, the density of voters largely follows the density of the population. Matching polling booths to colonies and localities is a fairly straightforward task. We suggest sampling polling booth areas with a probability based on risk of infection (not just number of voters), using readily-available characteristics.

The first level of selection of areas will be driven by ICMR guidelines, since rapid antibody-based blood test for COVID-19 is possible only in areas reporting clusters (containment zone). This will be restricted to the booths that are within a 3km radius of a confirmed cluster of cases. Within the booths that lie within the 28 sq. km. area, the choice can be prioritised by:

(a) The number of infections already recorded in the locality

(b) The density of the population

(c) The type of neighborhood (e.g., slum or planned)

It is very important to include existing information at the sampling stage, while keeping randomisation intact. This makes the data useable for further analysis to iteratively generate efficient testing, a valuable feature for analysis, during an ongoing pandemic with a scarcity of tests (see no. 4 below).

3 . Selecting People to Test:

A random sample of residents within a polling booth **is unlikely to maximise** the probability of finding individuals being infected. A combination of three strategies is suggested to find the infected population: (i) screening, (ii) local information, and (iii) local data collation.

Screening: One should go to a random set of households and ask if they fall into a series of "high risk" categories in the population based on a small set of screening questions (see Annexure). If the household is deemed to be high risk, then testing should be conducted (while retaining basic information on the houses that were not tested). The details of the proportion of people falling into these categories, from a random sample, should be recorded. This will allow us to estimate the infection rate among a high-risk population, as well as the extent to which they can be found in the population.

The screening questions relate to three categories: (a) potential for being infected, e.g., healthcare workers, police, emergency workers, etc. (b) potential for infecting others because of occupation, healthcare workers, police, and (c) potential for infecting others because of the extent of contact e.g. social workers, religious leaders, etc. If infection is detected in this group, it is difficult to trace contacts and isolate. Workers (men more likely to be working than women), should be tested with a higher probability than non-workers.

Local Information: If one infers from the scale of detected infections in Mumbai and Pune, where testing has occurred at a higher rate, it is likely that infection has spread beyond the small number of professions detailed above. The use of records from all local health clinics, hospitals, etc. about those who have sought remedy for symptoms associated with influenza like illness (ILI) will increase the chance of finding people who may be inflected with COVID-19, who were missed in a regime of limited testing.

Local Data Collation: In the highest risk areas, a number of people will have been tested, some positive, but mostly negative. Knowing how many such individuals are in the locality, and whether they came forward to be tested, will give some sense of the extent to which infection can be caught by self-detection, as opposed to the state information gathering mechanisms described above.

4. Iteration of Procedure and Real-Time Analysis.

The data should be collected digitally, collated, and analysed on daily basis to inform the next places to sample and visit. The logic is simple – as we get to know that regions of the city, or certain types of population, are of particular concern, we can define a better model of high-risk areas and individuals, which can be fed into the testing protocol on a day-to-day basis.

This requires rapid, high-quality data analysis. Towards this end, we suggest making all data public to the maximum extent possible in an anonymized form with geo-referencing or cluster identification to make use of skilled data analysts nationally and internationally to quickly analyse and offer feedback on the data – adopting a "crowding-in" approach to data analysis. This daily calibration will optimise precious medical and financial resources in identifying the infected population and spreading behavior.

A beginning can be made on the public database by making unit level data on the individuals already tested available in a public database in an anonymised manner in a model similar to the crowd sourced patient database **https://t.me/COVID19indiaops**. This information would form the core of the database mentioned above and additional data would be added as tests are conducted. States like Maharashtra (https://drive.google.com/file/d/1sYzTQM77cpBI2xs7cvDBvy0WJJsLRXGQw/view) are releasing more data and information than others, but are not crowding in analysis.

Asymptomatic testing

This can be further enhanced by convincing the ICMR to allow for testing of asymptomatic persons, given the nature of this virus (**in Iceland**, which has tested the largest percentage of its population of any country, **about half the persons infected by COVID-19 were asymptomatic**), and by allowing testing of blood samples from the general population, as in blood banks for the presence of COVID-19. This will provide a rough and ready way of measuring prevalence among the population (subject to the nature of blood donors vis-à-vis the general population).

Pooled Samples

Another innovation is to allow the use of pooled samples. This, of course, depends on the sensitivity of the test (whether the test will be able to detect the presence of COVID-19 if one positive sample is diluted with other negative samples). **If this is tested and found acceptable**, one can pool the tests for an area – a polling booth has about 300 to 400 households – which will increase the speed by which areas can be included/ excluded as well as minimise the number of tests required.

Concluding Thoughts

These are trying times for the entire world. In India, there are fears that the inability to catch the spread of coronavirus will overwhelm the medical system and state resources. But high-quality, flexible policy, built on iterative data collection, with public datasets will allow information and many data analysts to guide the optimal places and groups within which to conduct coronavirus testing. The procedures detailed above are natural for most applied and survey statisticians and can be implemented quickly. Using available data at each step to guide the locating of high-risk populations and iterating this process will yield an efficient use of resources in minimising the transmission of the virus.

ANNEXURE: SCREENING SURVEY

I. Basic Information about the Respondent

1)   Demographics
(i)   First Name
(ii)   Last Name
(iii)   Gender
(iv)   Marital Status
(v)   Date of Birth

2)   Identity and Contact information:
(i)   Voter ID Card Number (this is the database from which selected)
(ii)   Passport Number (if applicable):
(iii)   Mobile Number
(iv)   Mobile Number (Alternative - Emergency Contact)
(v)   Email Address

3)   Contact
(i)   On an average day, how many people would you say you interact with?
(ii)   On an average day, how many people would you say you shake hands with?
(iii)   On an average day, how many people would you say you hug?

4)   Do you have any of the pre-existing conditions?
(i)   Diabetes
(ii)   Hyper-tension
(iii)   Respiratory Problems
(iv)   Obesity
(v)   Cancer
(vi)   Smoking
(vii)   Others – please specify
(viii)   None of the above

5)   Have you experienced any of the following symptoms in the last two weeks?
(i)   Fever
(ii)   Flu/Runny Nose
(iii)   Dry Cough
(iv)   Sore Throat
(v)   Joint Pain
(vi)   Difficulty Breathing
(vii)   Diarrhea
(viii)   None of the above

6)   Have you been close to anyone showing any of the following symptoms: Fever, Flu or Cough?
(i)   Yes
(ii)   No
(iii)   Do not Know

7)   Have you been close to anyone who has been diagnosed with Coronavirus in the last two weeks?
(i)   Yes
(ii)   No
(iii)   Do not Know

II. Respondent Travel and Interaction History

8) **Have you recently (in the last month) travelled domestically (within your country)?**

(i)   Yes

(ii)   No

If yes, which city or cities did you travel to?

9) **Have you recently (in the last month) travelled internationally (outside your country)?**

(i)   Yes

(ii)   No

If yes, which country or countries did you travel to?

(i)	China	(vii)	Italy
(ii)	Saudi Arabia	(viii)	Spain
(iii)	Iran	(ix)	Germany
(iv)	United Arab Emirates (UAE)	(x)	France
(v)	United Kingdom (UK)	(xi)	Other
(vi)	United States of America (USA)		

10) **Have you been in contact with someone who has recently (in the last month) travelled domestically (within country)?**

(i)   Yes

(ii)   No

(iii)   Do not Know

If yes, which city or cities did that person travel to?

11) **Have you been in contact with someone who has recently (in the last month) travelled internationally (outside country)?**

(i)   Yes

(ii)   No

(iii)   Do not Know

If yes, which country or countries did that person travel to?

(i)	China	(vii)	Italy
(ii)	Saudi Arabia	(viii)	Spain
(iii)	Iran	(ix)	Germany
(iv)	United Arab Emirates (UAE)	(x)	France
(v)	United Kingdom (UK)	(xi)	Other
(vi)	United States of America (USA)		

III. Respondent Daily Commute and Household

12)   How do you travel during the day?

(i)   Public Transport

(ii)   Private Car

(iii)   Uber/Ola

(iv)   Motorbike

(v)   Rickshaw

(vi)   Bicycle

(vii)   Walk

How many hours do you spend during the day on your commute?

13)   How many people under the age of 20 are living in your household?

14)   How many people over the age of 60 are living in your household?

IV. Perception of COVID-19

15)   Have/Did you take(n) any preventive measures to prevent you and your family members from being infected by COVID-19 (Coronavirus)?

(i)   Yes

(ii)   No

If yes, what measures did you take?

a)   Washing hands with soap for 20 seconds

b)   Use of hand sanitizer

c)   Use of surgical face mask

d)   Use of N95 face mask

e)   Use of disinfectant at home and workplace

f)   Take anti-viral medicine

g)   Social Distancing

h)   Other

16)   Do you think 'social distancing' is important as a precautionary measure?

(i)   Yes

(ii)   No

(iii)   Do not Know

17)   Have you attended any events with more than 10 people in the last month?

(i)   Yes

(ii)   No

If yes, Area/City where you attended social gatherings