

Data to help manage pandemics in the Global South¹

Rohan Samarajiva²

rohan@lirneasia.net; +94777352361

27 March 2020

Contents

- Introduction2
- Exclusions.....2
- Data-based insights, by purpose3
- Optimal allocation of scarce resources and targeting of interventions3
- How big will it be?.....3
- Where will it hit?.....4
- Contact tracing.....6
- Enforcing quarantine7
- Building trust.....7
- Concluding comments9



¹ The research reported in this note includes work supported by the International Development Research Centre of Canada.

² With contributions from Lasantha Fernando, Yudhanjaya Wijeratne, Sriganesh Lokanathan, Nuwan Waidyanatha & Helani Galpaya.

Introduction

A pandemic is an epidemic that crosses national or continental boundaries. Inherently, the effects of pandemics are global. But because the health systems are weaker, and a greater proportion of the population is in the informal sector, the effects in terms of deaths and livelihoods may be stronger in the Global South. Therefore, novel applications of data analytics are of great value in the Global South. However, lower levels of datafication and network and smartphone penetration pose challenges.

Because diseases such as COVID-19 and Ebola are transmitted person-to-person, the effects are broader than people getting sick or dying. Intuitive responses may be anti-social and regressive.³ Prevention may involve certain actions that may disrupt social and economic activities. Prevention also requires identification of and effective isolation or treatment of persons who have been in contact with persons diagnosed as diseased. Given the possibility that asymptomatic individuals may still transmit the disease, various forms of mobility and contact restrictions may be applied to a relatively large number of persons. Behavioral change and restraint are central elements. Therefore, data, information and knowledge can assist in implementing efficient and effective responses. The intention of this note is to provide a tour d'horizon, rather than explore issues in depth. Where LIRNEasia has engaged in relevant research, it will be mentioned.

This tour d'horizon will place emphasis on the possible uses of data to help stop or slow the spread of the disease directly. It gives weight to what can be done in the short term.

Exclusions

Obviously, data can help the larger policy process by documenting the extent of the incidence of the disease and the rate of propagation. This can be done using official data,⁴ or other data.⁵ They will be critical in increasing the salience of the issue and prodding the decision-makers to act, or the various pressure groups to exert pressure. But they are not prioritized in this note.

Data can also support early detection of an emerging epidemic or pandemic. Ideally, this knowledge, if seen as credible, will lead to preventive measures such as travel restrictions and preparations such as stockpiling of medical supplies. Reportedly, a Canadian health monitoring platform which utilizes artificial intelligence (AI) tools detected and announced the COVID-19 outbreak in the Hubei Province in China nine days before the WHO announcement and six days before that by the US Centers for Disease Control and Prevention.⁶

The above example falls within the scope of syndromic surveillance. Syndromic surveillance is "indicator-based" surveillance. It is about finding anomalies against a baseline (or baselines with seasonal effects). For example, if an abnormal count of similar fever and respiratory symptoms are detected among an age group in a single location, that could serve as a trigger for epidemiologists to engage in deeper analysis. Given the focus is on finding deviations from the norm in massive sets of

³ Van Bavel, et. al. (2020). Using social and behavioural science to support COVID-19 pandemic response.

<https://psyarxiv.com/y38m9>

⁴ <https://blog.watchdog.paladinanalytics.com/tracking-the-spread-of-covid-19-in-sri-lanka-a-data-story/>

⁵ Volipelli, G. (2020). Hidden data is revealing the true scale of the coronavirus outbreak. Wired.

<https://www.wired.co.uk/article/coronavirus-spread-data>

⁶ Niiler, E. (2020 Jan 25). An AI epidemiologist sent the first warnings of the Wuhan Virus, Wired.

<https://www.wired.com/story/ai-epidemiologist-wuhan-public-health-warnings/>

data, the task is well suited for AI applications as has been demonstrated by the apparent success of BlueDot, which uses natural-language processing (NLP) and machine learning to discern patterns and anomalies in news reports in 65 languages. BlueDot uses airline data and reports of animal disease outbreaks, but not social media content. It does not have access to actual syndromic surveillance data from outpatient treatment centers and hospitals.

More than a decade ago LIRNEAsia tested the possibilities of using feature phones equipped with Java-based applets and drop-down menus in two districts in India and Sri Lanka to capture syndromic surveillance data in datafied form and using software that would today be described as narrow AI to identify deviations from the norm.⁷ For this approach to be sustainable over time, the work processes within hospitals and outpatient units would have to be fully computer-based so that that data used in syndromic surveillance is a by-product of the diagnosis and treatment, rather than an additional chore. In many cases, the data would be available for analysis even before the diagnosis is complete. However, the entire subject is outside the scope of this note because its implementation requires a long-term and systemic effort, primarily involving the medical establishment.

Data analytics and AI have transformed medical research, especially genomics.⁸ Indeed, the world would be in much greater difficulty today if not for this. The COVID-19 disease is estimated to have originated in the period December 5-23, 2019.⁹ The mapped genome was submitted by Chinese researchers to an open source repository on January 17, 2020.¹⁰ This speed could only have been possible because of advanced computing now pervading genomics.

Data-based insights, by purpose

Optimal allocation of scarce resources and targeting of interventions

How big will it be?

In the case of a rapidly growing epidemic policy makers need to be able to forecast what resources (i.e., how many ventilators, beds, etc.) will be needed, where and when. This requires the ability to forecast demand, namely the expected number of patients, classified by levels of care required and location.

Researchers seek to use time series models (auto regressive and ARIMA [Auto Regressive Integrated Moving Average] models) for forecasting. But in the early stages of the epidemic there are not enough data for time-series modelling, somewhat obviously. The alternative is to find epidemic curves that are like the target country/locale using K-means clustering. This proved promising, but with countries making different kinds of interventions the similarities dissolved. With the limited data, it proved difficult to get the time series to be stationary, which is required for modeling time series data. Another approach is using an SIR [Suspected, Infected, and Removed] /SEIR [Susceptible, Exposed, Infected, and

⁷ Gow, G., et al. (2010). Using mobile phones in a Real-Time Biosurveillance Program: Lessons from the frontlines in Sri Lanka and India. <https://ieeexplore.ieee.org/document/5514617>

⁸ Chivers, T. (2018). How big data is changing science, Mosaic. <https://mosaicscience.com/story/how-big-data-changing-science-algorithms-research-genomics/>

⁹ Zhang, C., & Wang, M. (2020). Origin time and epidemic dynamics of the 2019 novel coronavirus. *bioRxiv*.

¹⁰ Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, https://www.ncbi.nlm.nih.gov/nuccore/NC_045512

Resistant] model for which an app has been developed.¹¹ However, estimating model parameters remains a challenge.¹²

Modelling work has been hampered by difficulties of obtaining reliable data. In Sri Lanka, data from the Epidemiological Unit was not in adequately digitized form. The numbers tested and the number of patients treated at hospitals did not tally. Details were missing regarding the residence of a patient, whether it was a community transmission or an imported case, etc. A recently launched app, covidsl.com, has a good interface but doubts exist about the quality of the underlying data. There have also been publications on a correlation between the weather parameters and propagation of COVID-19, but again, quality of data and difficulties of aligning patient and weather data (which are reported for different geographical areas) exist.

Where will it hit?

In the case of diseases such as dengue, the emphasis has been on predicting locations where the disease is most likely to emerge. Contact tracing has not been of importance, because the disease is transmitted by mosquitoes who have been infected, rather than directly by humans. Instead, the practice has been for public-health officials to take prophylactic measures targeted to the micro locations (such as home, office or school) where the infected person may have spent significant time. Prediction models developed by researchers¹³ can be used to allocate scarce medical resources efficiently or to target preventive measures.

Instead of the after-the-fact targeting of micro locations as in the case of dengue, here the objective is to identify larger locales such as cities, districts or provinces before the fact. Models that predict human movements at particular time may have proven useful, for example in containing the original outbreak in Wuhan by identifying the main locations of travel from Wuhan during the Chinese New Year.¹⁴ These models would be based on historical travel patterns and may not be fully accurate for a specific year, but it would still improve resource allocation and targeting of interventions. If, for various reasons, travel cannot be stopped, this would allow the optimal placement of limited thermal scanners and other devices.

Each disease has symptoms. In the case of COVID-19 elevated temperature is an indicator. If all or a significant proportion of the population has connected temperature sensors on their bodies or are using thermometers capable of automatically reporting temperatures, that could be a source of data that can be analyzed for insights.¹⁵ Sensors could be handed out for no charge, as is being done to some extent

¹¹ <https://alhill.shinyapps.io/COVID19seir/>

¹² See also gabgoh.github.io/COVID/index.html which implements SEIR with some tuning for COVID-19 parameters.

¹³ Wesolowski, A., et al. (2015). Impact of human mobility on the emergence of dengue epidemics in Pakistan. PNAS September 22, 2015 112 (38) 11887-11892. <https://doi.org/10.1073/pnas.1504964112>; Fernando, L., Lokanathan, S., Perera, A. S., Ghouse, A., & Tissera, H. A. (2017). Improving Disease Outbreak Forecasting Models for efficient targeting of Public Health Resources. Proceedings of 12th Communications Policy Research South (CPRSOUTH) Conference; see also, an approach that does not use mobile data:

<https://www.csiro.au/en/News/News-releases/2019/New-tool-to-track-human-infectious-diseases-in-Australia>

¹⁴ For an example of such work using Sri Lankan mobile network data, see <https://lirneasia.net/2017/11/using-call-data-records-analyze-event-attendance/>

¹⁵ McNeil, D.G., Jr. (2020 March 18). Can Smart Thermometers Track the Spread of the Coronavirus? New York Times. <https://www.nytimes.com/2020/03/18/health/coronavirus-fever-thermometers.html#click=https://t.co/MoS9TUXAZe>

in the case of COVID-19. Devices such as Fitbit can communicate far more than just temperature. Alternatively, apps can be loaded on to smartphones, or smartphones with apps installed could be given to a random sample or a class of people. The reports can be pseudonymized, in which case the impacts on individual privacy will be minimal. Clusters of high-temperature individuals may be interpreted as an outbreak of disease, and test kits and other resources moved to that location. But this may have impacts on what is described as collective or group privacy.¹⁶

The large populations of developing countries, gaps in wireless coverage and cost make it unlikely that specialized human-body sensor networks will be a practical option at the present time. But if they are used, there will be broad privacy concerns that will have to be addressed. The “big data” or the transaction-generated data that will be most effective in all countries of the Global South is Mobile Network Big Data (MNDB) comprising Call Detail Record (CDR) and Visitor Location Registry (VLR) data. The former depends on calls, texts and Internet searches being done by the user (or the persons connecting with them). The latter is a form of things communicating with things, also known as Internet of Things or IoT. They yield insights on social networks and on physical mobility (where the phone and its carrier has been at what time). Network data (obtained from operators) provides the best coverage, independently of terminal devices and GPS.¹⁷

CDR and VLR data can yield rich insights, about the population but also about individuals. When pseudonymized data is used, there are few concerns about personal data.¹⁸ For the models to be of optimal use, it would be necessary to use near real-time data. Pseudonymizing data requires time, especially if data from multiple operators are used, as should be. Depending on the weight placed on speed, personally identifiable information (PII) may or may not be involved. If the data are used without proper pseudonymization, it may require exemptions from data protection laws.

However, the predictions will identify locations that are likely to be clusters. The “group privacy” of the people living in those locations will be affected in that they may be subject to various forms of interventions and constraints. But this is a necessary and desirable identification. Unless groups who are likely to be infected or are likely to infect others are identified in time, people will die.

A prejudice against actions based on group attributes would frustrate efforts to improve the functioning of society in systematic, evidence-based ways. For example, it is routine to associate various characteristics or behaviors with persons living in geographical areas (e.g., in poverty mapping), by age group and gender and so on. It is considered desirable to “target” various policy measures to specific groups and indeed to improve the targeting by various means. Without group identification it will be impossible for modern societies to function. This is possibly the reason why safeguards against group identification do not currently exist in law and are not likely to exist in the future.¹⁹

¹⁶ Taylor, L.; Floridi, L.; van der Sloot, B. (2016). Group privacy: New challenges of data technologies. Springer.

¹⁷ Tirone, J., Seal, T.; Drozdiak, N. (2020 March 18). Location Data to Gauge Lockdowns Tests Europe’s Love of Privacy. Bloomberg. <https://www.bloomberg.com/news/articles/2020-03-18/austria-italy-join-push-to-use-mobile-data-to-gauge-lockdown>

¹⁸ Primarily the concern is about re-identification of the pseudonymized data by using other large datasets where the same individuals are represented.

¹⁹ Samarajiva, R. & Lokanathan, S. (2016). Using behavioral big data for public purposes: Exploring frontier issues of an emerging policy arena. LIRNEasia & Open Society Foundation. <http://lirneasia.net/wp-content/uploads/2013/09/NVF-LIRNEasia-report-v8-160201.pdf>

Contact tracing

Quick control of diseases that are transmitted person-to-person depends on prompt and complete identification of all who have come into contact with the person who has tested positive for the disease.²⁰ WHO has developed an application called Go.Data to assist with contact tracing. The tool enables “case investigation, contact follow-up, visualization of chains of transmission including secure data exchange and is designed for flexibility in the field.”²¹ Singapore and South Korea appear to have mastered this in the case of COVID-19.²² It appears at present that in Singapore the contact tracing is primarily done using traditional police methods, rather than the use of MNBD or other data. The traditional methods appear to be working well because people either trust the government or believe that the retribution likely to follow a lack of cooperation is bound to be severe. In South Korea, officials use security camera footage, credit card records, even GPS data from vehicles and mobile phones.²³ It appears that in Taiwan, massive data sets such as health insurance and immigration data bases are being integrated and analyzed.²⁴

The Singapore government has developed a community smartphone application called TraceTogether.²⁵ The app sends short-distance Bluetooth signals among phones of participating proximate users. The app estimates the distance between the app users and the duration of such encounters. Encounter records are encrypted and stored locally on phones for 21 days, the incubation period of the COVID-19 virus. If a user were to be diagnosed as infected, and with user consent, this data is then shared with authorized government officials who would access the data and obtain the mobile numbers of the user’s contacts.²⁶ Sandy Pentland of MIT is planning to launch a Beta version of a similar app that is GDPR compliant in a week.²⁷ The WHO is also developing a similar application.²⁸

²⁰ The steps involved in contact tracing including contact identification, contact listing, and contact follow are set out in WHO guidelines at <https://www.who.int/features/qa/contact-tracing/en/>

²¹ <https://www.who.int/godata>

²² Vaswani, K. (2019 March 19). Coronavirus: The detectives racing to contain the virus in Singapore, BBC. <https://www.bbc.com/news/world-asia-51866102>; Ng, Y. et al. (2020). Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore — January 2–February 29, 2020. <https://www.cdc.gov/mmwr/volumes/69/wr/mm6911e1.htm>;

²³ Fisher, M.; Choe, S. (2020, March 23). How South Korea flattened the curve, New York Times, https://www.nytimes.com/2020/03/23/world/asia/coronavirus-south-korea-flatten-curve.html?campaign_id=2&emc=edit_th_200324&instance_id=16983&nl=todaysheadlines®i_id=9770121&segment_id=22692&user_id=e4707ae60680263c08fd207686a8f6e1

²⁴ Beech, H. (2020 March 17). Tracking the Coronavirus: How Crowded Asian Cities Tackled an Epidemic, New York Times. <https://www.nytimes.com/2020/03/17/world/asia/coronavirus-singapore-hong-kong-taiwan.html>

²⁵ <https://www.channelnewsasia.com/news/singapore/covid19-trace-together-mobile-app-contact-tracing-coronavirus-12560616>

²⁶ <https://mothership.sg/2020/03/tracetogether-installed-open-source/>

²⁷ <http://safepaths.mit.edu>

²⁸ <https://spectrum.ieee.org/the-human-os/biomedical/devices/who-official-coronavirus-app-waze-covid19>

In situations where the identified patient has difficulty recalling all the persons who have been in contact or is uncooperative, investigators would be tempted to seek evidence from CCTV cameras,²⁹ or records from telecom operators. Legal authority has been given to Shin Bet, the Israeli internal security agency to use MNBD to trace contacts.³⁰ It has been claimed without too much evidence that telecom operators in Taiwan and South Korea have provided governments with individual-level data to track contacts.³¹ The narrative is much stronger in terms of state surveillance in the case of China.³² Not only data from the network itself but also from nearby WiFi networks and Bluetooth beacons are said to be used. What is intriguing is that universities and other entities appear to be collaborating with the UK government to build similar capabilities into apps that are to be installed on willing users' phones.

Enforcing quarantine

When large numbers must be quarantined and their movements controlled, it is natural to think of technology. Geo-fencing is a technique that has been used to generate alerts or to disable certain devices when they move outside a defined perimeter.

Data from mobile operators can be used to track the movements of persons ordered to self-isolate or enter quarantine. Depending on the sophistication of the mobile operators' systems, the location resolution in this data would be finer than just the location of a base station. Urban areas with higher base station density can provide for a more localized estimation of a user. However, these can be further augmented by continuously triangulating and tracking the movements of those isolated/quarantined.

Hong Kong has implemented a system based on wrist bands, smartphones and actions that have to be taken by the person in quarantine (e.g., taking a photograph and sending it within a tight time frame, in Singapore).³³ What is feasible in developed economies such as Singapore and Hong Kong where almost universal availability of smartphones may be assumed, may not be possible in countries of the Global South.

Building trust

It has been found that trust by users in the organization that is implementing some change contributes to success. In many countries in the Global South, the state lacks legitimacy. The public does not trust it to keep its word. Conspiracy theories abound as do anecdotes about rules being bent for the benefit of the powerful. In these conditions, introducing technological means to help overworked officials with necessary actions such as contact tracing and enforcing quarantine is likely to prove difficult. In countries such as South Korea and Singapore where the population has experience of earlier epidemics such as SARS and MERS and where the state has actually performed well in taking people out of poverty and providing public services, it appears that there is a greater receptiveness to technological solutions

²⁹ <https://www.hindustantimes.com/india-news/coronavirus-covid-chaos-from-lucknow-to-lutyens/story-EzgmjcfP8VPIUe0zbpbdRn.html>

³⁰ <https://www.nytimes.com/2020/03/16/world/middleeast/israel-coronavirus-cellphone-tracking.html?referringSource=articleShare>

³¹ <https://www.reuters.com/article/us-health-coronavirus-europe-telecoms/european-mobile-operators-share-data-for-coronavirus-fight-idUSKBN2152C2>

³² <https://www.nytimes.com/2020/03/19/us/coronavirus-location-tracking.html>

³³ <https://qz.com/1822215/hong-kong-uses-tracking-wristbands-for-coronavirus-quarantine/>

such as proximity-recording apps. Just because they work in such societies, there is no guarantee of acceptance in poor countries of the Global South.

In many countries, infection equals stigma. The notice that had been issued by Air India asking for the cessation of discriminatory and hostile acts against its employees by neighbors in housing societies³⁴ exemplifies the problem. Healthcare professionals are also being discriminated against.³⁵ Here, it appears that ill-informed and untrusting residents of housing societies assume that all who travel abroad are carriers of infection. They are driven by fear and lack the information to differentiate between greater and lesser risks. This leads to unreasonable discrimination against entire classes of people.

To condemn discrimination against doctors, nurses or airline personnel is not to say that the risks of infection through them is not significant. Addressing the fear that drives the overbroad, anti-social behavior requires first that the fear be understood and acknowledged. Instances exist of public policy recognizing such fears and seeking to assuage them, even at the cost of violating principles such as redemption and privacy. Many rich countries subscribing to liberal values maintain registries of persons who have committed “sex offences” and completed their punishment. In the USA, these registries are open to the public.³⁶

One solution that does not violate rights is to seek to convince the fearful members of the public of the unreasonableness of their behaviors and the public good that is served by supporting workers such as healthcare and transport professionals who are engaged in providing essential services.³⁷ The efficacy of this solution in low-trust societies is patchy at best, but there is no alternative to trying. However, it may be worth complementing it with an additional measure, that of transparency: giving people credible information about testing procedures and the precautions being taken to ensure that frontline personnel performing essential tasks are protected from infection.

But this cannot be done for every single member of society, only for specific classes of frontline employees. Controversial though it may be the dashboard that can be seen by anyone about every single COVID-19 case in Singapore may be seen as an attempt to provide a general solution.³⁸ In addition to a network diagram that shows all the infection clusters and cases, the dashboard also discloses information about infected persons, short of the name. That means that information such as address, workplace, age, etc. are published along with details of contacts and medical condition. On the face, this level of disclosure appears overbroad and violative of privacy as commonly understood in many societies.

Were a similar dashboard be provided in a low-trust society, the results may be different. Most likely, the conversation will shift to how various adjustments have been made for the rich and the powerful or about how certain patients have been kept out of the records altogether. However challenging it is, it

³⁴ <https://timesofindia.indiatimes.com/business/india-business/air-india-crew-being-ostracised-by-neighbours-housing-societies-for-operating-flights-to-covid-19-countries/articleshow/74761456.cms>

³⁵ <https://krdo.com/news/national-world/2020/03/25/doctors-evicted-from-their-homes-in-india-as-fear-spreads-amid-coronavirus-lockdown/>

³⁶ https://en.wikipedia.org/wiki/Sex_offender_registry

³⁷ <https://krdo.com/news/national-world/2020/03/25/doctors-evicted-from-their-homes-in-india-as-fear-spreads-amid-coronavirus-lockdown/>

³⁸ https://co.vid19.sg/cases?fbclid=IwAR34H9bDUJ1IUo_0SuBypkdd9R-JpucBPe5xbblpYN0N2VaQoC56O2fhBjg

seems that new approaches to the problem of building trust is critical to the management of pandemics. There is merit in looking at the Singapore dashboard in the context of trust-building, ensuring that the necessary safeguards are established, etc. Simply staying with the low-trust equilibrium is not an option.

Concluding comments

This short tour d’horizon provided an overview of actions that may be taken using “big data” as well as “small data” to provide solution that would contribute to improved management of pandemics. Emphasis was placed on solutions appropriate for the context of the Global South. It is possible to quickly adapt some technological solutions developed in rich countries for contact tracing and quarantine enforcement. The modeling for early detection and for resource allocation using big data is more challenging, in that issues of adequately datafied records and intra-state frictions about releasing or sharing data have to be addressed. The larger question of building or rebuilding trust using technological or other means also has to be given priority.