Gathering personal data for smart contact tracing – How to encourage (and not discourage) participation

Ole Jann, CERGE-EI — March 25, 2020.

Summary:

- South Korea and Singapore show that the disruption of anti-virus measures can be minimized through intelligent data collection and contact tracing. Not only did the two countries conduct many tests, but they also used the information they gathered to trace the spread of the infection as quickly as possible through personal interviews, medical data, mobile phone data, credit card transaction data and CCTV recordings. Strict measures were then imposed on those connected to the infection chain, which allowed for softer widespread measures with smaller effects on the economy. This strategy is also possible in the Czech Republic, but it requires the intelligent design of rules and incentives. Each country has only one attempt to get this right.
- Governments should aim to incentivize citizens' cooperation – but even more importantly, need to make sure not to create incentives for non-cooperation. Trying to force the revelation of personal, sensitive data for contact tracing may instead incentivize hiding symptoms or recent contacts.
- Collecting and storing personal data for contact tracing is a more sensitive topic in Europe than in some East Asian countries, as there is less tradition of community enforcement and a higher distrust of the state. Strong data protection is therefore key to a successful tracing strategy. Even a low level of non-compliance, not to mention a culture of being "clever" about the rules, could render the contact tracing system ineffective.
- Personal data gathered for contact tracing must be kept beyond the reach of law enforcement, tax investigators, the civil court system and researchers who do not work directly on combating the virus. Data collection and storage must be clearly time-limited.
- Lack of data protection and a resulting unwillingness to cooperate with contact tracing efforts could lead to further breakouts with enormous loss of life and damages to the health of citizens and the economy.

Policies and initiatives around the world

Despite having early outbreaks of Covid-19, South Korea and Singapore saw very slow growth of infection numbers and deaths. Both countries have done lots and lots of tests, but have also used these to follow infection chains as quickly as possible, using personal interviews, medical data (especially visits to medical facilities), cell phone data, credit card transaction data, and CCTV recording. Some of this information is made available to the public, such as in this map. They have also imposed *strict* measures on those in any way connected to the chain, which allows them to impose *softer* measures on the general population. In sum, this allows them to slow the spread of the virus sufficiently with a minimal impact on the economy and society. This seems our best hope of restarting normal life and the economy in the coming months.

There are now many initiatives around the world to improve data gathering and contact tracing, especially with the help of smartphone apps, including from places such as MIT and the University of Oxford. Closer to home there is an initiative by Covid19cz to roll out "targeted, data-driven predictive testing". (The latter claims to be GDPR compliant.) Any such efforts must necessarily rely on detailed data about health status, movements and social contacts.

How to set the right incentives and get useful information

Gathering the data that allows for efficient differentiation between groups relies critically on the cooperation of citizens. Even if providing information is mandatory, in many situations it will be impossible to thoroughly check people's statements and it is therefore critical that infected patients as well as the general population are motivated to cooperate in contact tracing efforts. This is obvious in the case of interviewing infected people about their contacts – but South Korean authorities, for example, also found it hard to match phone numbers to patients without their cooperation.

Any effort to gather detailed and personalized data for the purpose of tracking contacts and infections may create perverse incentives to hide symptoms or conceal contacts. This may be a larger problem in some European countries than in East Asia, since there is less tradition of community enforcement and strong distrust of the state. Good policies must hence strike a balance between effective measures and the adverse incentives they may create. Economic theory offers three main lessons:

1. Most people will be willing to cooperate since they understand the necessity of fighting the virus. But some incentives and checks might help to convince them to be truthful, where this is possible without creating perverse incentives. If cell phones or credit cards are used outside the home during mandatory quarantine (as they seem to have been in almost half of all cases), this could be reason for a police check-up.
2. Some incentives can be counterproductive. Small fines for lying or breaking quarantine can make it look like a trivial, everyday infraction, while if there are no fines the moral imperative is stronger.[1] Excessive punishment for breaking quarantine could lead people to hush up their symptoms rather than being quarantined.
3. Many people will have ample incentives not to cooperate in government data gathering. That is why the data should be gathered in a *separate, well-defined and time-limited* effort. The following paragraphs will discuss this point in more detail.

Data protection could save lives

It is not hard to imagine a cheating husband who does not mention all the places he has visited, a businessman who does not want to reveal his customers because he evades taxes, or the owner of a strip club who wants to protect his clients or staff. Any of these could lead to another outbreak that is only detected once lots of people have fallen ill. (In South Korea, one non-cooperative patient is believed to have been responsible for their outbreak growing from 30 to more than 8000 cases.) The data gathering needs to be value-neutral, and needs to

[1] This is sometimes called the "Israeli daycare effect" in behavioral economics – in an Israeli daycare, introducing fines for late pick-up increased the frequency of this behavior, cf. Gneezy and Rustichini (2000).

reassure the participants that their data will remain private and they will not be hurt if they help trace the spread of the virus.

In particular, the data collected

(i) should be kept separate from other government data wherever possible,
(ii) should be in a format that is hard to connect to other government data,
(iii) should be anonymized wherever possible (e.g. each person's data could be saved under a pseudonym, with the database linking name and pseudonym kept separately and only used in extraordinary cases),
(iv) should not be available for any other purpose than contact tracing and learning about infection frequency,
(v) should be beyond the reach of law enforcement, tax investigators, the court system (unless someone is accused of misreporting this data) and researchers not directly working on fighting the virus, and
(vi) should have firm and clear rules about when it will be deleted.
(vii) All these rules should, wherever possible, be given legal status. They should also be announced by government leaders to make sure that they are well known and that they are seen as official and intentional government policy.

The Czech government has already taken the sensible approach of not publishing detailed data about infections to avoid stigmatization of individuals and groups. Handling the highly personalized and private data involved in contact tracing requires an even more careful approach.

Economic research has shown that if people are concerned about how their data will be used, the information we learn from them can be useless.[2] Far from limiting the effectiveness of contact tracing, strong data protections may therefore be essential to its success. There is only one shot to get this right. Even if just a few people see an incentive not to cooperate, or (worse) a general culture of non-cooperation and "being clever" emerges (as it has with tax collection), such behavior could open the door for another outbreak that could cost many lives.

References

- Daughety, A. F. and Reinganum, J. F. (2010). Public goods, social pressure, and the choice between privacy and publicity. *American Economic Journal: Microeconomics* 2(2), 191–221.
- Gneezy, U., and Rustichini, A. (2000) A Fine is a Price. *Journal of Legal Studies*, 29(1), 1-17.
- Jann, O., and Schottmüller, C. (2020). An informational theory of privacy. *Economic Journal*, 130(625), 93-124,

[2] Cf. Daughety and Reinganum (2010) and Jann and Schottmüller (2020).