The need for privacy with public digital contact tracing during the COVID-19 pandemic

Lancet Digital Health 2020
Published Online
June 2, 2020
https://doi.org/10.1016/S2589-7500(20)30133-3

Digital contact tracing applications represent a powerful yet controversial strategy to combat the COVID-19 pandemic. Manual contact tracing has important challenges, not limited to recall bias and delays in communicating with high-risk contacts.[1] Digital technologies are already increasingly used in the context of health-care delivery and clinical trials.[2] Due to the considerable strain on public health institutions, digital contact tracing through mobile phones is being used or explored in a growing number of countries despite concerns raised over individual privacy and state surveillance.[3]

Mobile phone-enabled digital contact tracing colocalises individuals in time and space through the use of GPS, Bluetooth, or other such technologies. Google and Apple have promised to provide frameworks for how to use their technologies for contact tracing.[4] A digital contact trail can be created when individuals who have downloaded such applications come into physical proximity. Machine-learning strategies[5] can improve on simple binary contact tracing systems by providing methods to calculate quantifiable individual risk of acquiring COVID-19 depending on specific features such as distance and duration of interaction, self-reported comorbidities, demographics, and the presence of any symptoms in each individual in an interaction. As an individual's risk level for acquiring COVID-19 increases, various behavioural messages can be delivered quickly to enable the individual to take appropriate, measured action. These multiple advantages have the potential to establish rapid epidemiological control of the pandemic.[6]

Despite the potential advantages, most of the applications in use or under consideration have an impact on individual privacy that democratic societies would normally consider to be unacceptably high. In a free and democratic society, there are major concerns regarding privacy. The UK, Australia, Singapore,[3] South Korea, and other countries have deployed such tools (using binary variables of contact, not scalar risk probabilities for risk of infection); however, these applications have come under scrutiny relating to the ability of governments and other groups to access personal information.[7] Public trust in the use of these applications is paramount because widespread adoption of these technologies is needed to be effective in curbing viral transmission. Indiscriminate collection of personal information, chronic privacy breaches, and lax attitudes towards individual privacy in the private sector have eroded public trust in digital technologies. Moreover, tracing applications raise the spectre of generalised state surveillance in the face of the pandemic, with potentially

Panel: Recommendations for a privacy-protecting approach to digital contact tracing

Consent

- Download, installation, and use of the application must be entirely voluntary, and users must be able to uninstall the application at will
- There must be express consent for all collection, use, and disclosure of personal information (ie, users might choose to share some data and not others, such as official test results or to feed a machine-learning model)
- Individuals must be able to opt-in or opt-out of data sharing. This includes consent to download the application, turn on location services, receive notifications, and share COVID-19 test results

Oversight

- A non-partisan independent oversight committee with representatives from legal, health, machine-learning, and privacy experts should be established to oversee ongoing development of the application, its information ecosystem, and data governance
- Importantly, public representatives must be included in this oversight committee

Virtual data acquisition

- No identifiable information regarding digital contact trails or personal health information that an individual enters on the application should be shared with other application users or public, private, and governmental agencies
- Individual geolocation data should not be stored on a central server and should pass through a rigourous obfuscation protocol to reduce their information content to the bare minimum required for epidemiological and machine-learning modelling
- Pseudonymised data should be used to inform machine-learning models, and only these data should be stored centrally on a protected server
- Only non-identifiable aggregated data should be shared with public health institutions
- The source code of the application and the algorithms used should be made accessible for public scrutiny
- Personal identifiable information should be deleted from the device once the pandemic is over

Informed decision making

- User preferences should drive end-to-end experience
- User comprehension should be prioritised and verified rather than assumed
- User psychosocial wellbeing should be promoted
- User empowerment to protect themselves and others should be maximised
- User inclusivity should acknowledge the diversity of user needs in dimensions such as gender, race, education, and rural *vs* urban location

devastating consequences if democratic societies learn to accept such an intrusion on civil liberties.[8] Therefore, to counteract both negative perceptions and genuine threats, a privacy-protecting approach must be central in the development of such a contact tracing application.

Several strategies can be leveraged to increase and maintain the public trust with such applications (panel). Express consent at each step of data sharing is crucial and must be meaningful, not buried within lengthy privacy policies or vague language agreements, and includes express consent to anonymously share COVID-19 test results. No identifiable data should be shared with any public institution or private enterprise. Pseudonymised or aggregate data can be adequately used to develop machine-learning and epidemiological models and inform public policy. Otherwise data should be kept encrypted on users' devices and inaccessible to public authorities or private interests. The tracing application itself can propagate alerts to high-risk contacts and can recommend that users voluntarily contact health authorities where relevant, thereby assisting markedly in contact tracing while minimising the potential for state surveillance, snooping, or vigilantism.

The granular non-identifying information used to train machine-learning models generally contains sufficient detail to re-identify individuals when correlated with other sources of data. This is why an independent, non-partisan trust or similar fiduciary structure must be established to protect and control access to these data, and manage the application and its ongoing development. The source code for the application and the privacy protocols used should be publicly available. Individuals must be able to make independent informed choices based on recommendations released from the application rather than using coercive or penalising strategies. An application self-destruction strategy should be used so that once the pandemic is over, all application-related personal data is deleted from participants' phones and deleted from the machine-learning server, leaving for further research, only de-identified, aggregated, and statistical data, or artificial data generated from the epidemiological model.

For **details on the application** see https://arxiv.org/abs/2005.08502

The approach presented here advocates that consent must be explicating for users to download the application, transmit COVID-19 test results, and share data for research. Recent projections suggest that at least 56% of a country's population would need to be using the application to ensure maximal chance of epidemiological control of the COVID-19 pandemic.[9] There is a tension between mandating use of the application versus having a consent-based approach that we are advocating. In the face of such tension, the trade-off between individual civil rights and the need for population-level control of the COVID-19 pandemic comes to the forefront. Trust in the application by individuals is pivotal for such applications to have population-level benefit. We would suggest that advocating an approach that emphasises consent and prevents any central public or private authority from accessing identifiable data would embolden more individuals to download the application, thereby optimising the population-level benefit. Various designs are currently in place with regard to strategies for identifying contacts, the types of notifications that are received, and the use of centralised versus decentralised approaches.[4,10] One question that arises in a system that emphasises a consent-based, opt-in approach, is that among individuals who do not receive a notification, does the absence of the notification imply the absence of contacts with other individuals with a COVID-19 infection or that other users are not consenting to share data? The absence of notifications might create a false sense of security in the user of the application or can cause frustration if a user presumes that others are not sharing information. This limitation with such opt-in applications emphasises the need for broad public outreach and education to optimise the number of users who download the application and consent to share data.

Leveraging digital contact tracing technologies can change the course of the COVID-19 pandemic. Such technologies must robustly support democratic principles of privacy to maintain public trust and to enable individuals to make informed choices to help combat the pandemic.

We are part of a team developing a COVID-19 risk awareness application in Canada. YWY reports funding from the Toronto COVID-19 Action Initiative. DP, BS, and SK received funding support from Mila to assist in application development. AS reports grants from Fonds de la Recherche en Sante du Quebec—Junior 1 clinician scientist programme and Bristol-Myers Squibb-Pfizer, personal fees from Novartis and AstraZeneca, grants and personal fees from Roche Diagnostics and Boehringer-Ingelheim, and funding from the McGill Interdisciplinary Initiative in Infection and Immunity (Mi4), outside the submitted work. All other authors declare no competing interests.

*Yoshua Bengio, Richard Janda, Yun William Yu, Daphne Ippolito, Max Jarvie, Dan Pilat, Brooke Struck, Sekoul Krastev, *Abhinav Sharma*
abhinav.sharma@mcgill.ca

Montreal Institute for Learning Algorithms, Université de Montréal, Montreal, QC, Canada (YB); Faculty of Law (RJ) and McGill University Health Centre Research Institute (AS), McGill University, Montreal, QC H4A 3J1, Canada; Department of Computer and Mathematical Sciences, University of Toronto, Toronto, ON, Canada (YWY); Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA (DI); Borden Ladner Gervais, Montreal, QC, Canada (MJ); and The Decision Lab, Montreal, QC, Canada (DP, BS, SK)

1 Sun K, Viboud C. Impact of contact tracing on SARS-CoV-2 transmission. *Lancet Infect Dis* 2020: published online April 27. https://doi.org/10.1016/S1473-3099(20)30357-1.

2 Sharma A, Harrington RA, McClellan MB, et al. Using digital health technology to better generate evidence and deliver evidence-based care *J Am Coll Cardiol* 2018; **71:** 2680–90.

3 Servick K. Cellphone tracking could help stem the spread of coronavirus. Is privacy the price? *Science* 2020; published online March 22. DOI:10.1126/science.abb8296.

4 Apple, Google. Privacy-preserving contact tracing. https://www.apple.com/covid19/contacttracing (accessed May 11, 2020).

5 Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521:** 436–44.

6 Ferretti L, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 2020; **368:** eabb6936.

7 Lin L, Martin TW. How coronavirus is eroding privacy. April 15, 2020. https://www.wsj.com/articles/coronavirus-paves-way-for-new-age-of-digital-surveillance-11586963028 (accessed May 11, 2020).

8 Harari YN. Yuval Noah Harari: the world after coronavirus. March 20, 2020. https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75 (accessed May 11, 2020).

9 Hinch R, Probert W, Nurtay A, et al. Effective configurations of a digital contact tracing app: a report to NHSX. April 16, 2020. https://cdn.theconversation.com/static_files/files/1009/Report_-_Effective_App_Configurations.pdf?1587531217 (accessed May 20, 2020).

10 Alsdurf H, Bengio Y, Deleu T, et al. COVI white paper. May 18, 2020. https://arxiv.org/abs/2005.08502 (accessed May 20, 2020).