

A pilot study using financial transactions' spatial information to define high-risk neighborhoods and distribution pattern of COVID-19

Esmaeil Mohammadi¹, Mehrdad Azmin¹, Nima Fattahi¹, Erfan Ghasemi¹, Sina Azadnajafabad¹, Negar Rezaei^{1,2}, Mohammad-Mahdi Rashidi¹, Mohammad Keykhaei¹, Hossein Zokaei^{1,2}, Nazila Rezaei¹, Rosa Haghshenas¹, Farzad Kaveh³, Erfan Pakatchian¹, Hamidreza Jamshidi⁴ and Farshad Farzadfar^{1,2} 

Abstract

Background: Development of surveillance systems based on big data sources with spatial information is necessitated more than ever during this pandemic. Here, we present our pilot results of a new technique for the incorporation of spatial information of transactions and a vital registry of COVID-19 to evaluate the disease spread.

Methods: We merged two databases of laboratory-confirmed national COVID-19 registry of Iran and financial transactions of point-of-sale devices from February to March 2020 as our training data sources. Spatial information was used for the visualization of maps and movements of sick individuals. We used the point-of-sale devices-related guild to check for the dynamics of financial transactions and effectiveness of quarantines.

Findings: In the study period, 174,428 confirmed cases were in the COVID-19 registry with accompanying transactions information. In total, 13,924,982 financial transactions were performed by them, with a mean of 1.2 per day for each person. All guilds had a decreasing pattern of “risky” transactions except for grocery stores and pharmacies. The latter showed a decreasing pattern by impose of lockdowns. Different cities were the hotspot of disease transmission as many “high-risk” transactions were performed in them, among which Tehran (mainly its central neighborhoods) and southern cities of Lake Urmia predominated. Lockdowns indicated that the disease gradually became less transmissible.

Interpretation: Financial transactions can be readily used for epidemics surveillance. Semi real-time results of such iterations can be informative for policy makers, guild owners, and general population to prepare safer commuting and merchandise spaces.

Keywords

COVID-19, transactions, surveillance, epidemics, pandemic

Submission date: 11 August 2021; Acceptance date: 10 January 2022

¹Non-Communicable Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran

²Endocrinology and Metabolism Research Center, Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran

³Center for Communicable Diseases Control, Ministry of Health & Medical Education, Tehran, Iran

⁴Research Institute for Endocrine Sciences, School of Medicine, Department of Pharmacology, Shahid Beheshti University of Medical Sciences, Tehran, Iran

Corresponding author:

Farshad Farzadfar, Non-Communicable Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, No. 10, Al-e-Ahmad and Chamran Highway Intersection, Tehran 1411713137 Iran.
 Email: f-farzadfar@tums.ac.ir

Research in context

Evidence before this study

Big data and vital registries have the potential to improve the quality and speed of surveillance in the context of epidemics. We searched PubMed and Google Scholar from February 2019 (as when COVID-19 emerged) to July 2021 using search terms related to the aim of study [“big data”, “machine learning”, “artificial intelligence”, “data science” in Boolean association with “COVID-19”, and “surveillance”] and retrospectively expanded the search for other conditions. Search yielded mostly experimental studies on the use of big data sources such as public transport, surveillance cameras to supervise the virus while none utilizing the financial transactions. While there were small experiments on the use of credit card transactions for surveying the infected cases, this was not used over time for the virus distribution behavior and assess the efficacy of implemented public health interventions. We, therefore, evaluated the effectiveness of this data source for integration with vital registries like COVID-19 aiming for epidemiologic surveillance of disease.

Added value of this study

Our study evaluated the effectiveness of a big data source (financial transactions) with both spatial and temporal information to be integrated and used in epidemiologic surveillance of a fast-spreading condition. This method and source have not been previously used in large scale and can be successfully integrated and automated for epidemiologic supervision purposes.

Implications of all the available evidence

As we are entering the era of “datafication”, large and available data sources of different commodities, such as financial databases, can be integrated into other vital ones to improve our temporal and spatial accuracy of disease and epidemic supervision. This analysis could add information about the pattern of virus propagation and measures that are relatable to the future pandemics.

Introduction

Acute respiratory syndrome related to the new coronavirus was declared as a pandemic by the World Health Organization (WHO) on March 2020.^{1,2} It has led to the death of millions of people, economic crises, and markets shutdown. Surveillance approaches have been implemented around the world for better control of the disease spread and prevention.³ As it has been well emphasized “Amidst this crisis, one institutionalized response promises a modicum

of certainty: surveillance”, the importance of supervision measures is crystal clear.⁴ After a while, Iranian governments faced financial tensions, and passed acts to restart the economic activities in a controlled fashion, a matter known as “controlled social distancing”^{5,6}. Efforts were made to make this notion “intelligent” and surveyed as much as possible to trade-off the disease horizontal distribution in some part to support the economy. Supervisions were carried out by use of health system load indicators and statistics.^{7,8} On the other hand, considering Iran as a fast-senescing population with rapid epidemiologic transition, burden of predisposing mortality risk factors of COVID-19 is demandingly high and surveillance measures should be meticulously utilized to assess the efficacy of lockdowns and passed act.^{9–12} One attempt was the utilization of spatial information of banking system transactions to uncover the safety of city locations, working and marketing areas, and detection of hotspots of disease transmission. Infographics and reports were made available to the public and governors to discover high-risk places for disease transmission in a semi real-time fashion.² In this study, we report our experience on the combination of data sources of this communicable disease for surveillance purposes.

Methods and material

This study was approved by the Endocrinology and Metabolism Research Institute and Tehran University of Medical Sciences ethics committees (IR.TUMS.EMRI.REC.1399.039). Data were retrieved from two independent sources: (1) Ministry of Health and Medical Educations’ registry of COVID-19 reverse transcriptase polymerase chain reaction (RT-PCR) confirmed cases and (2) Records of Central Bank of Iran. Data of the current report are from February to March 2020 (i.e. phase one of the epidemic in Iran). This method was performed biweekly to this date and reports are provided to health system authorities. COVID-19 registry is the sum up of all screening tests performed around the country and includes everyone diagnosed with a lab test. It is important to note that in this pilot study, positive cases detected in the first phase of COVID-19 epidemic were enrolled, and due to the unavailability of diagnosis kits in many areas, there remained many undiagnosed cases. Spatial information of performed transactions on point-of-sale (POS) terminals was retrieved in the mentioned duration. Considering that in crowded spaces the risk of contagion increases, only POS devices were chosen. Another advantage of using these devices is that their exact location is registered. Besides the location of each financial transaction, the guild related to each device (e.g. grocery stores or gas stations) and the national IDs were fetched. Because of the pandemic and risk of transmission, guilds were made to not accept paper cash,

and considering it as an untraceable transaction, we only chose payments on POS devices. A person can hold many cards registered to any given service provider bank, but all transaction information is automatically and simultaneously transferred to the Central Bank and summed up to the national ID. Central Bank is the upper-level organization of all other service provider banks in Iran. All transaction data are simultaneously and automatically recorded in Central Bank's database. National ID is a unique code given to every Iranian national after their birth. To protect the confidentiality of identified data, only one member of the data analysis team (EG) was assigned for the preprocessing of datasets on a single unlinked computer that we use to keep restricted information. Afterward, the preprocessed data were recoded and unidentified for further analysis. Data were cleaned for missing values, mispronounced inputs, or duplicates. Two data sources were merged by the recoded IDs. The density of transactions performed by test-positive individuals in different cities was used to find higher-risk locations. Movement of "high-risk" individuals between cities was also investigated to assess the between-city distribution of disease. As the Persian New Year holidays (Nowruz, March 21 until 2 weeks) was included in the study period, the dynamic of transaction could be utilized for evaluating people's commuting, incompliance to the universal lockdowns (from early March), and how effective preventions were implemented. Several subsequent steps were carried out for the visualization and creation of infographics. Details of methodology for each infographic and graphic are provided in the related figure caption. To evaluate the effectiveness of lockdowns and limitations on between-city travels, we used a network graph that was composed of edges as transactions performed by a test-positive person in a city [node] outside the location their debit card was issued. All analyses were performed in Python (Python Language Reference, version 3.6., Available at: www.python.org).

Results

In total, 174,428 laboratory-confirmed COVID-19 patients' information were included, being responsible for 13 924,982 transactions on POS devices. Of them, 97,932 (56.2%) were male patients and the mean age of sample was 52.9 ($\pm SD$: 19.3) years. Also, 145,639 (83.5%) of patients were first diagnosed in an inpatient setting while for 190 (0.1%) patients, the setting was missing. Considering the symptom of cough as a high potent distributor of virus, 84,923 patients had active tussis at the time of the first diagnosis. The mean number of transactions performed was 116,041 "high-risk" transactions in each day. In other words, each sick individual during the study period did 1.2 daily "high-risk" transactions.

Over time, the total number of transactions gradually decreased and the pattern was intensified on the second

week of March as the New Year holidays approached (Figure 1A and 1B). This pattern was similarly detected in commercial guilds, too, with varying intensities. Although it was increasing at first for grocery stores and pharmacies, they turned downward just before the holidays. Inside the grocery stores, those dedicated to vegetables and fruit trade were also found to be the most dangerous "super spreader" locations. Opposed to our impression, the number of transactions performed for purchases related to Internet providers and mobile devices was stagnant, relating to unpreparedness of Iranian infrastructure during the early stages of epidemic.

During the first week of March, as the peak of the first phase of the epidemic, most northern and western areas and cities were densely infected with the virus, based on the number of "high-risk" transactions. In the next 2 weeks, limitations and lockdowns were implemented and enforced, leading to "dilution" of dense areas while populated cities of Tehran, Karaj, and southern region of Lake Urmia remained hotspots (Figure 2A, B, and C). Considering neighborhoods of large cities, word plots were created as infograms to represent the high-risk locations. In Tehran, central populated areas hosted more high-risk transactions (Figure 3). Efforts were made to withhold the distribution of disease between cities, but the movement of sick individuals based on the location of their transactions indicated active commuting of people not only in their provincial areas but also long-distance travels to the capital (Tehran) or other crowded cities (Figure 4). In other words, there were plenty of debit card transactions performed by sick individuals in far cities other than the city of issuance, indicating nonadherence to lockdowns during the first phase of the epidemic. Even many movements were detected to small touristic cities and islands.

Discussion

The main point of this study is that spatial and geographical information data sources from non-health sectors, like financial transactions, can be successfully integrated into health datasets for surveillance purposes especially in the context of fast-spreading epidemics. Enforced lockdowns were related to a decrease in the density and crowdedness of shops and markets. Besides, unpreparedness of Internet-based platforms withheld the lockdowns in their bests.

Surveillance systems are the cornerstone of any controlling and prevention strategy and are a crucial sector of public health organizations of each nation.¹³ It is clearly emphasized that real-time measures of disease burden and dispersion are required for placement of feasible and yet effective interventions and increase the general awareness and knowledge of community.^{14,15} One major drawback of health care systems in developing countries is the lack of an efficient surveillance system.¹³ Many electric and

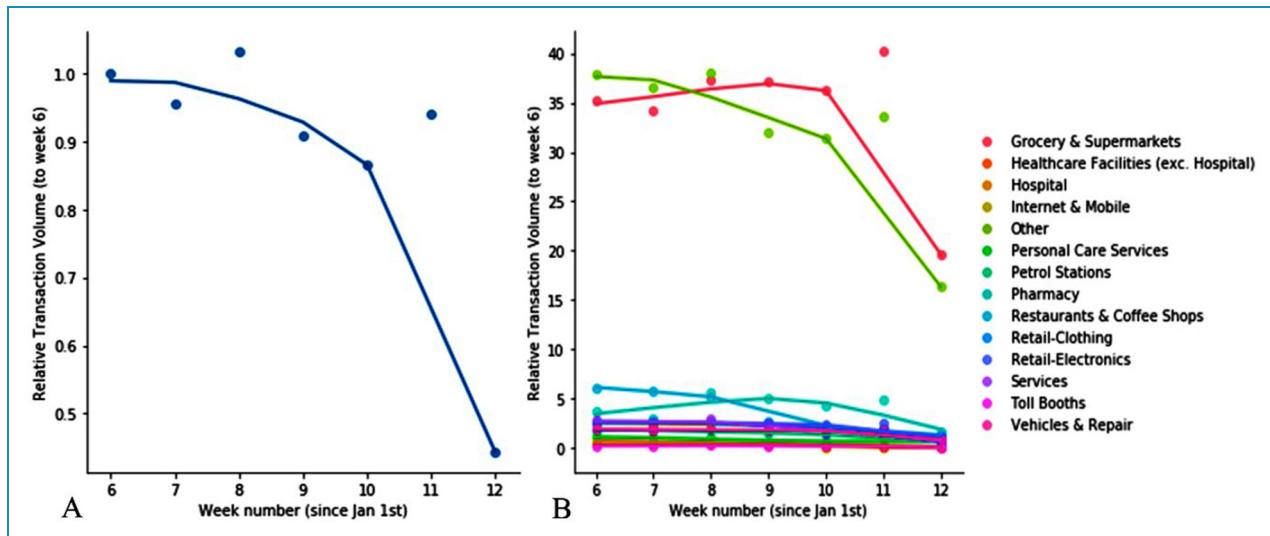


Figure 1. Loess curves representing the weekly pattern of transactions over the first phase of COVID-19 epidemic in Iran. (A) Based on total transactions performed relative to the amount performed on January 6 and (B) in each guild. Y-axis of the right-hand panel represents the percentage of transaction volume for the guild in that particular week. Points indicate the transaction load without any modeling, adjustment, or imputation.

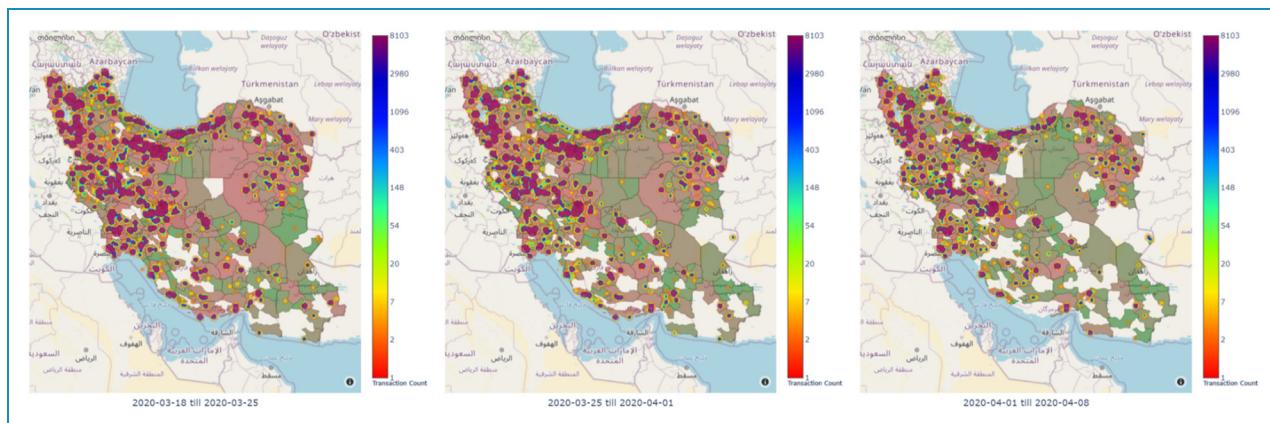


Figure 2. Distribution pattern and density of “high-risk” transactions in cities of Iran around March 21, the Persian New Year holidays. (A) Two weeks before the start of universal lockdowns, (B) the 2 weeks period of holidays, and (C) 2 weeks afterward.

digital surveillance systems are in action in different countries, in which many bioinformatics platforms are utilized to interactively elaborate real-time multi-measure data sources. Such platforms indeed necessitate large investments, although the outcome is much greater.¹⁶

Other countries’ experience on financial transactions and payment measures indicate that financial data are real-time readily available high-resolution information that can be relied on for surveillance purposes especially in epidemics and lockdowns.¹⁷⁻¹⁹ A Spanish experiment on 2.1 billion transactions indicates that the mobility of individuals has changed during COVID-19 pandemic. They elaborate a divergence in the mobility amount of low-wage and wealthier populations and incompliance of such amenable groups to lockdown laws during weekdays.¹⁷ A report



Figure 3. Word plot illustrating the density of high-risk financial transactions in different neighborhoods of Tehran in late March 2020.

from French banks showcases a sudden decrease in the consumption of money in their withheld accounts, while a rebound is being detected after a while, representing the clinging of people to lockdowns at the early stages of epidemic and slow incompliance to the restrictive rules. Another interesting finding was that wealthy deciles were more likely to save money and lower-wage groups were more likely to face debt.²⁰ Several other studies have pointed out the dynamic of transactions and assets during the COVID-19 pandemic with similar annotations,^{21,22–26} but none have reported the possible role of transactions

and banking system data for disease transmission surveillance which the WHO has elaborated its importance.²⁷

Big data and artificial intelligence can be effectively used for the detection of sick individual's mobility and implementation of preventive strategies.²⁷ Chinese authorities have extensively used big data sources for disease surveillance and prevention, among the integrated sources are transportation system databases, mobile phones, and social media can be found.²⁷ South Korea has also inspired COVID-19 Smart Management System (COVID-19 SMS) that integrates security credit card transactions,

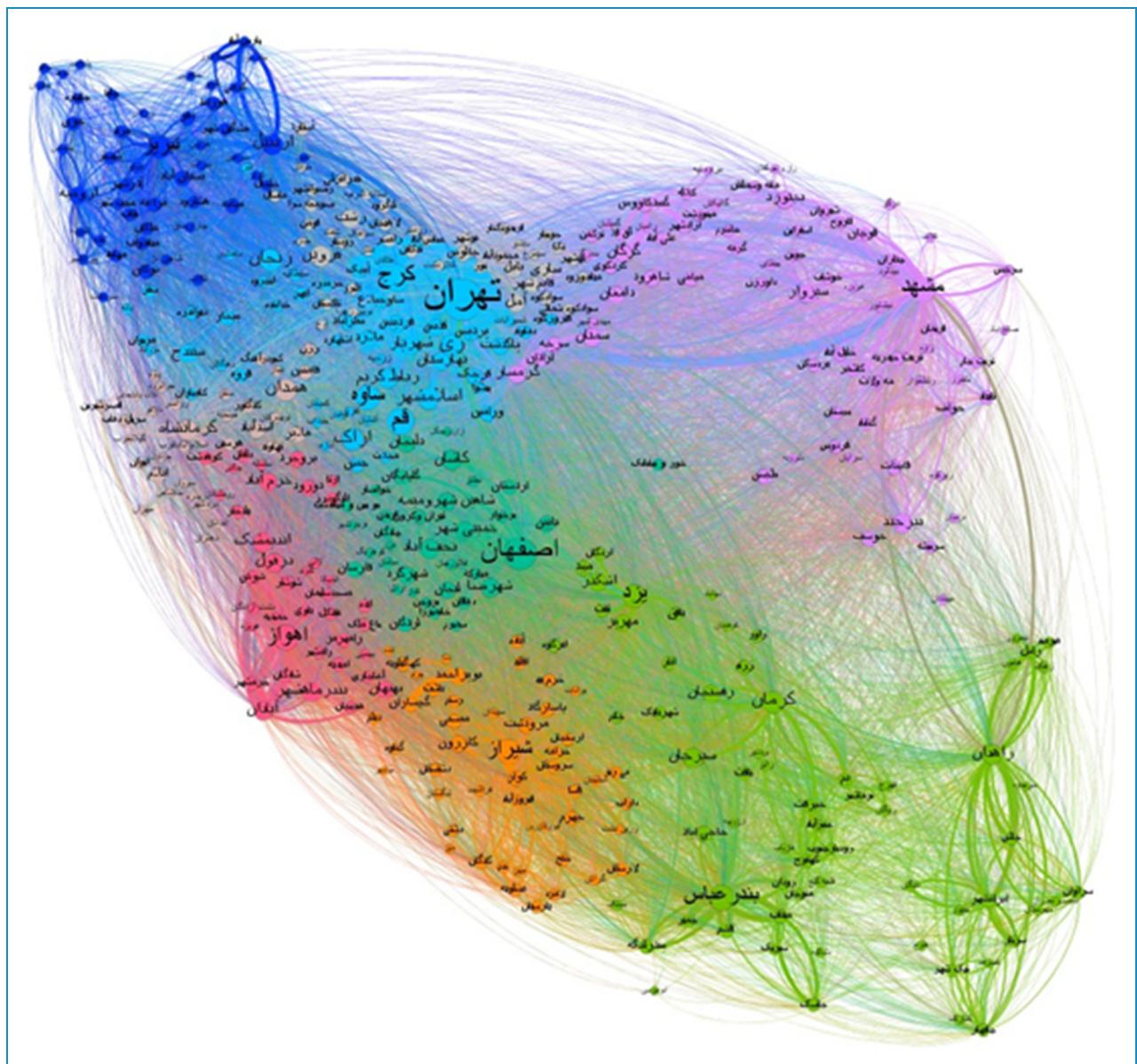


Figure 4. Network graph representing the traveling and dispersion of genetically positive cases over different locations during the first phase of the epidemic in Iran. Edges represent transactions performed by individuals in a location other than their home city during the first phase of the epidemic in Iran. Larger nodes refer to cities that hosted higher risky transactions from “external” individuals. Colors indicate the geographical clustering of cities into northern, eastern, etc., areas.

smartphones location, and security camera records to trace sick individual's movements. Singapore's experience on mobile apps and Bluetooth technology or Taiwanese use of cellular data for restriction and controlling sick people was also a success.^{28,29} Similar other experiments were also detected in literature from the United States, United Kingdom, and Japan to track the movements of sick individuals.³⁰⁻³² To date, these countries have successfully contained the disease. Although privacy annotations may restrain the use of such actions to be performed in other countries.

Limitation, strengths, and future perspective

We believe that the current approach can be a feasible technique for surveillance of disease spread and rate of incompliance of infected persons. But there are several shortcomings for this approach that needs to be mentioned to find better solutions. First of all, this system is semi real time. Other surveillance systems such as identity detecting surveillance cameras or smartphone applications can reveal a timelier output. On the other hand, this approach cannot notify individuals about their accommodation through a risky location and presence of an actively contagious person in their close adjacency; or prevent sick individuals from getting mobilized.

One main limitation of current work is that due to constraints, real-time data retrieval was not possible, although the lag between registration of data and emergence of a sick individual was 3 days at most. Moreover, COVID-19 registry of Ministry of Health is not only restricted to laboratory-confirmed cases and encloses many are diagnosed with radiographic requests and history taking, while for the homogeneity of findings, we only included genetically confirmed samples. On the other hand, a novel integration of a spatial source of information from the banking system with a disease repository was first used in this study as far as we could acknowledge. Unavailability of data for healthy and unaffected individuals was another limitation to compare and assess the efficacy of lockdowns and other preventive measures, a matter that should be sought in future works. Using the guilds and categorization data of bank accounts enabled us to understand high-risk marketing places and inform policy makers and owners to employ safety measures. Implementation of innovative iteration strategies in real-time platforms makes more use of such efforts. Additionally, a combination of multiple spatial data information (e.g. roadside cameras and surveillance systems on automobile plates, transactions, cell phones, etc.) could lead to more accurate results.

Conclusion

Nowadays, we have access to the increasing amount of temporal and spatial data. Best practice is whenever these data

sources are utilized to mitigate the disease spread by restricting individuals from closely contacting each other.³³ Innovative approaches and combinations of nowadays vast sources of data can be utilized in the context of newly emerged crises and epidemics. Inter-sectorial collaboration of organizations and cooperation are vital for fruitful preventive strategy implementation. Trial and error is an inseparable component of innovation and invention. Support for these small-scale efforts rather than de novo investment in unbacked commodities is crucial for the achievement of larger end results. We believe geographical and spatial information of the banking system can be successfully utilized and integrated with other vital data sources for the aim of disease control and other purposes.

Acknowledgments: The authors are grateful for the support and diligence of Endocrinology and Metabolism Research Institute of Tehran University of Medical Sciences, Ministry of Health and Medical Education, and Central Bank of Iran. We would like to thank the Shifa Pharmed Industrial Group for their scientific support.

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Contributorship: EM analyzed the data and prepared the manuscript. MA, NF, and EG curated and analyzed the data and approved the final version of manuscript. SA, NegarR, MMR, and MK did the literature search and approved the final version of manuscript. HZ, NazilaR, and RH interpreted the results and approved the final version of manuscript. FK and EP approved the final version of manuscript. HJ scientifically consulted and approved the final version of manuscript. FF supervised the study, approved the final version, and is the corresponding author of this work.

Ethical approval: This study was approved by Endocrinology and Metabolism Research Institute of Tehran University of Medical Sciences ethics committee (IR.TUMS.EMRI.REC. 1399.039).

Funding: The author(s) received no financial support for the research, authorship and/or publication of this article.

Data sharing: The data used in this manuscript cannot be shared due to possession of identified information.

Informed consent: Not applicable, because this article does not contain any studies with human or animal subjects.

ORCID iD: Farshad Farzadfar  <https://orcid.org/0000-0001-8288-4046>

Trial registration: Not applicable, because this article does not contain any clinical trials.

References

1. Smith AC, Thomas E, Snoswell CL, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). *J Telemed Telecare* 2020; 26: 309–313.
2. Azadnajafabad S, Saeedi Moghaddam S, Rezaei N, et al. A report on statistics of an online self-screening platform for COVID-19 and its effectiveness in Iran. *Int J Health Policy Manag* 2021. doi:10.34172/ijhpm.2020.252
3. Wang S, Ding S and Xiong L. A new system for surveillance and digital contact tracing for COVID-19: spatiotemporal reporting over network and GPS. *JMIR Mhealth Uhealth* 2020; 8: e19457.
4. French M and Monahan T. Dis-ease surveillance: how might surveillance studies address COVID-19? *Surveill Soc* 2020; 18: 1–11.
5. Samadi AH, Owjimehr S and Nezhad Halafi Z. The cross-impact between financial markets, covid-19 pandemic, and economic sanctions: the case of Iran. *Journal of Policy Modeling* 2021; 43(1): 34–55. doi:10.1016/j.jpolmod.2020.08.001
6. Saki M, Ghanbari MK, Behzadifar M, et al. The impact of the social distancing policy on COVID-19 incidence cases and deaths in Iran from February 2020 to January 2021: Insights from an interrupted time series analysis. *Yale J Biol Med* 2021; 94(1): 13–21.
7. Arab-Mazar Z, Sah R, Rabaan AA, et al. Mapping the incidence of the COVID-19 hotspot in Iran—implications for travellers. *Travel Med Infect Dis* 2020; 34: 101630. doi:10.1016/j.tmaid.2020.101630
8. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, et al. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. *JMIR public health and surveillance* 2020; 6: e18828.
9. Aminorroaya A, Fattahi N, Azadnajafabad S, et al. Burden of non-communicable diseases in Iran: past, present, and future. *Journal of Diabetes & Metabolic Disorders* 2020: 1–7. doi:10.1007/s40200-020-00669-z
10. Azadnajafabad S, Mohammadi E, Aminorroaya A, et al. Non-communicable diseases' risk factors in Iran; a review of the present status and action plans. *Journal of Diabetes & Metabolic Disorders* 2021: 1–9. doi:10.1007/s40200-020-00709-8
11. Fattahi N, Azadnajafabad S, Mohammadi E, et al. Geographical, gender and age inequalities in non-communicable diseases both at national and provincial levels in Iran. *Journal of Diabetes & Metabolic Disorders* 2021. doi:10.1007/s40200-020-00713-y
12. Mohammadi E, Aminorroaya A, Fattahi N, et al. Epidemiologic pattern of cancers in Iran; current knowledge and future perspective. *Journal of Diabetes & Metabolic Disorders* 2020; 20(1): 825–829. doi:10.1007/s40200-020-00654-6
13. Ibrahim NK. Epidemiologic surveillance for controlling covid-19 pandemic: types, challenges and implications. *J Infect Public Health* 2020; 13: 1630–1638.
14. Ibrahim N and Al Bar H. Surveillance of childhood vaccine-preventable diseases at health facilities in jeddah. *Saudi Arabia. EMHJ-Eastern Mediterranean Health Journal* 2009 2009; 15: 532–543.
15. Fan C, Liu L, Guo W, et al. Prediction of epidemic spread of the 2019 novel coronavirus driven by spring festival transportation in China: a population-based study. *Int J Environ Res Public Health* 2020; 17.
16. Mahmood S, Hasan K, Carras MC, et al. Global preparedness against COVID-19: we must leverage the power of digital health. *JMIR Public Health and Surveillance* 2020; 6: e18980.
17. Carvalho VM, Hansen S, Ortiz A, et al. Tracking the COVID-19 crisis with high-resolution transaction data. 2020.
18. Baker SR, Farrokhnia RA, Meyer S, et al. How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic. *The Review of Asset Pricing Studies* 2020; 10: 834–862.
19. Campos-Vazquez RM and Esquivel G. Consumption and geographic mobility in pandemic times. Evidence from Mexico. *Rev Econ Househ* 2021; 19: 353–371.
20. Bounie D, Camara Y, Fize E, et al. Consumption dynamics in the COVID crisis: real time insights from French transaction bank data. *Covid Economics* 2020; 59: 1–39.
21. Chetty R, Friedman JN, Hendren N, et al. *The economic impacts of COVID-19: evidence from a new public database built using private sector data*. National Bureau of Economic Research, 2020 <https://www.nber.org/papers/w27431>.
22. Hacioglu Hoke S, Käning DR and Surico P. Consumption in the time of Covid-19: Evidence from UK transaction data. 2020.
23. Chen H, Qian W and Wen Q. The impact of the COVID-19 pandemic on consumption: Learning from high frequency transaction data. Available at SSRN 3568574 2020.
24. Andersen AL, Hansen ET, Johannessen N, et al. Consumer responses to the COVID-19 crisis: Evidence from bank account transaction data. Available at SSRN 3609814 2020.
25. Bounie D, Camara Y and Galbraith JW. Consumers' Mobility, Expenditure and Online-Offline Substitution Response to COVID-19: Evidence from French Transaction Data. Available at SSRN 3588373 2020.
26. Dunn A, Hood K and Driessens A. Measuring the effects of the COVID-19 pandemic on consumer spending using card transaction data. *US Bureau of Economic Analysis Working Paper WP2020-5* 2020.
27. World Health Organization. *Surveillance strategies for COVID-19 human infection: interim guidance*, 10 May 2020. World Health Organization, 2020. <https://apps.who.int/iris/handle/10665/332051>
28. Hui M. Singapore Wants All Its Citizens to Download Contact Tracing Apps to Fight the Coronavirus. 2020. <https://qz.com/1842200/singapore-wants-everyone-todownload-ad-covid-19-contact-tracing-apps/> (accessed June 2021).
29. Hui M. How Taiwan is Tracking 55,000 People Under Home Quarantine in Real Time. 2020. <https://qz.com/1825997/taiwan-phone-tracking-system-monitors-55000-under-coronavirus-quarantine/>.
30. Couture V, Dingel JI, Green A, et al. JUE insight: measuring movement and social contact with smartphone data: a real-time application to COVID-19. *Journal of Urban Economics*. 2021; 103328.

31. Thiemo F. Subsidising the spread of COVID-19: Evidence from the UK'S Eat-Out-to-Help-Out Scheme. *The Economic Journal* 2021; ueab074. doi:10.1093/ej/ueab074
32. Redding SJ, Glaeser EL and Gorback C. How Much Does COVID-19 Increase with Mobility? Evidence from New York and Four Other US Cities. 2020. <https://www.nber.org/papers/w27519>
33. Masuda N and Holme P. Predicting and controlling infectious disease epidemics using temporal networks. *F1000Prime Rep* 2013; 5: 6.