

A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys

To the Editor — In December 2019, a novel coronavirus was isolated, after a cluster of patients in China were diagnosed with pneumonia of unknown cause¹. This new isolate was named ‘SARS-CoV-2’ and is the cause of the disease COVID-19. The virus has led to an ongoing outbreak and an unprecedented international health crisis. The number of infected people is rapidly increasing globally² and most probably is a vast underestimation of the real number of patients worldwide, as infected people are contagious even when minimally symptomatic or asymptomatic³. The spread of the disease has presented an extreme challenge to the international community, and policy-makers from different countries have each chosen different strategies, depending on the local spread of the virus, healthcare-system resources, economic and political factors, public adherence, and their perception of the situation.

Coronavirus infection spreads in clusters, and early identification of these clusters is critical for slowing down the spread of the virus. Here we propose that daily population-wide surveys that assess the development of symptoms caused by the virus could serve as a strategic and valuable tool for identifying such clusters and informing epidemiologists, public-health officials and policymakers. We show preliminary results from an Israeli survey of a cumulative number of over 74,000 responses and call for additional countries to join an international consortium to extend this concept in order to develop predictive models. We expect such data will allow the following: faster detection of spreading zones and patients; acquisition of a current snapshot of the number of people in each area who have developed symptoms; prediction of future spreading zones several days before an outbreak occurs; and evaluation of the effectiveness of the various social-distancing measures taken and their contribution to reducing the number of symptomatic people. This information could provide a valuable tool for decision-makers in those areas in which strengthening of social-distancing measures is needed and those in which such measures can be relieved. Preliminary analysis shows that in neighborhoods with a confirmed

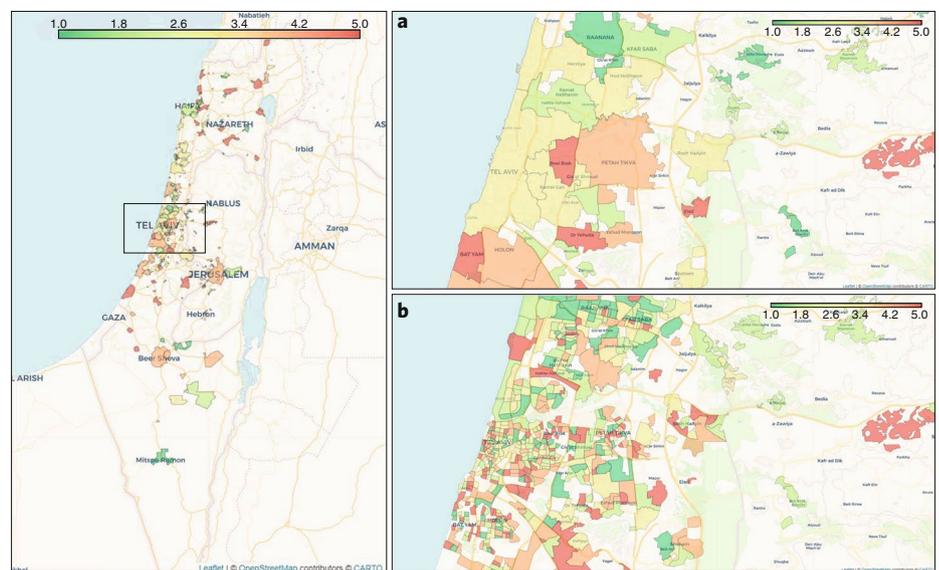


Fig. 1 | Average COVID-19-associated symptoms region map. City municipal regions with at least 30 completed surveys and neighborhoods with at least 10 completed surveys are shown. The color of each region indicates a category defined by the average symptoms ratio, calculated by averaging the reported symptoms rate by responses in that city or neighborhood. The values were divided into five categories, and the color of each region indicates its associated category, from green (low symptom rate) to red (high symptom rate) (key). **a**, Area of Tel-Aviv and Gush-Dan with city regions. **b**, Area of Tel-Aviv and Gush-Dan with neighborhood regions. Map data are copyrighted by OpenStreetMap contributors and are available from <https://www.openstreetmap.org>. Publ. note: Springer Nature is neutral about jurisdictional claims in maps. Credit: Leaflet | OpenStreetMap contributors | CARTO

patient history of COVID-19, more people report experiencing COVID-19-associated symptoms, which demonstrates the potential utility of our approach for the detection of outbreaks. Researchers from other countries, including the USA, UK, Canada, Switzerland, Germany, and several others are working on similar projects, such as the COVID Symptom Tracker in the UK⁴. We call with urgency for other countries to join our international consortium⁵, and to share methods and data collected from these daily, simple, one-minute surveys.

In Israel, the first infection of COVID-19 was confirmed on 21 February 2020, and in response, the Israeli Ministry of Health (MOH) instructed people who returned to Israel from specific countries in which COVID-19 was spreading to go into a 14-day home isolation. Since then, Israel has gradually imposed several additional

measures (Extended Data Fig. 1): on 9 March, the 14-day home isolation was extended to people arriving from anywhere of international origin, and those who were in close contact with a patient with confirmed COVID-19 were instructed similarly. Symptomatic people were instructed to stay home for 2 days after symptom resolution⁶. On 11 March, gatherings were limited to a maximum of 100 people; this was further restricted to 10 people on 15 March. On 19 March, a national state of emergency was declared in the country. On 20 March, the first death of an Israeli citizen from COVID-19 occurred.

One of the main challenges of the current pandemic so far has been disease detection and diagnosis. Although the gold standard for the diagnosis of COVID-19 is detection of the virus by a real-time PCR testing⁷, current resource and policy limitations

Table 1 | Characteristics of questionnaire responses

Characteristic (mean (s.d.) or %)	All (n = 74,256)	No in-home isolation (n = 71,249) (95.95%)	In-home isolation (n = 3,007) (4.05%)	Adults (n = 69,386) (93.44%)	Children (under 18 years of age) (n = 4,870) (6.56%)
Age (years)	45.38 (18.21)	45.54 (18.22)	41.58 (17.57)	47.73 (16.41)	11.95 (5.19)
Male	34,575 (46.56%)	33,085 (46.44%)	1,490 (49.55%)	32,108 (46.27%)	2,467 (50.66%)
Smoking history (previously smoked or currently smoking)	27,003 (36.36%)	25,990 (36.48%)	1,013 (33.69%)	26,814 (38.64%)	189 (3.88%)
Presence of a chronic medical conditions	15,102 (20.34%)	14,622 (20.52%)	480 (15.96%)	14,850 (21.4%)	252 (5.17%)
Symptoms					
No symptoms (feel good)	61,999 (83.49%)	59,592 (83.64%)	2,407 (80.05%)	58,191 (83.87%)	3,808 (78.19%)
Body temperature	36.5 °C (0.45)	36.49 °C (0.45)	36.53 °C (0.48)	36.49 °C (0.43)	36.62 °C (0.65)
Body temperature ≥ 38 °C	227 (0.31%)	196 (0.28%)	31 (1.03%)	130 (0.19%)	97 (1.99%)
Nausea and vomiting	433 (0.58%)	407 (0.57%)	26 (0.86%)	373 (0.54%)	60 (1.23%)
Myalgia (muscle pain)	2,359 (3.18%)	2,229 (3.13%)	130 (4.32%)	2,231 (3.22%)	128 (2.63%)
Rhinorrhea or nasal congestion	9,807 (13.21%)	9,384 (13.17%)	423 (14.07%)	8,804 (12.69%)	1,003 (20.6%)
Fatigue	3,014 (4.06%)	2,845 (3.99%)	169 (5.62%)	2,804 (4.04%)	210 (4.31%)
Shortness of breath	1,430 (1.93%)	1,331 (1.87%)	99 (3.29%)	1,343 (1.94%)	87 (1.79%)
Cough	10,089 (13.59%)	9,597 (13.47%)	492 (16.36%)	9,232 (13.31%)	857 (17.6%)
Diarrhea	1,286 (1.73%)	1,198 (1.68%)	88 (2.93%)	1,145 (1.65%)	141 (2.9%)
Symptoms ratio					
Symptoms ratio	0.05 (0.11)	0.05 (0.11)	0.06 (0.14)	0.05 (0.11)	0.05 (0.11)

The symptoms ratio (bottom row) was calculated as the number of reported symptoms divided by the number of symptoms in the predefined Israeli MOH symptoms list.

in many countries restrict the amount of testing that can be performed. The number of tests per day is increasing; however, not enough tests are being performed to provide a nationwide view of the spread of the virus, particularly as the Israeli MOH guidelines are to test only people who were in close contact with a person with confirmed COVID-19.

To obtain a real-time nationwide view of symptoms across the entire population, and since testing the entire population is not feasible, we developed a simple one-minute online questionnaire aimed at early and temporal detection of geographic clusters in which the virus is spreading. The survey was posted online (<https://coronaisrael.org/>) on 14 March, and participants were asked to fill it out on a daily basis and separately for each family member, including members who are unable to fill it out independently (e.g., children and older people). So that potential privacy issues that might occur can be avoided, our survey is filled out anonymously, and access to the data is restricted to only study investigators.

The survey contains questions on age, sex, geographic location (city and street), isolation status and smoking habits. Participants also report whether they are experiencing symptoms commonly described in patients with COVID-19 by

healthcare professionals, on the basis of the existing literature⁸. Several other symptoms that are less common in patients with COVID-19 but are more common in other infectious diseases are also included to better identify possible patients with COVID-19. The initial symptoms included cough, fatigue, myalgia (muscle pain), shortness of breath, rhinorrhea or nasal congestion, diarrhea and nausea or vomiting. Additional symptoms, including type of cough (with or without sputum), sore throat, headache, chills, confusion and loss of taste and/or smell sensation, were added in a later version. Participants also report about existing chronic health conditions and are asked to report their daily body temperature (Extended Data Fig. 2 presents the most recent version of the survey).

Given that reports on the clinical characteristics of patients with COVID-19 are only starting to emerge, we defined an initial basic measure we called the 'symptoms ratio' using symptoms that were predefined by the Israeli MOH and are commonly reported by patients with COVID-19⁸. Symptoms assessed were shortness of breath, fatigue, cough, muscle pains and fever (body temperature above 38 °C). For participants younger than 18 years of age, nausea and/or vomiting was also included in the ratio calculation.

For each participant, the symptoms ratio was calculated as the number of reported symptoms divided by the number of symptoms in our predefined list (number of reported symptoms / 6, for participants 18 years of age or less; number of reported symptoms / 5, for participants over 18 years of age). We plan to refine this list of symptoms as more clinical information is accrued. By associating participants with an area corresponding to their address, we created a color map of Israel by the aggregated symptoms ratio in each neighborhood (Fig. 1).

The questionnaire was first distributed online on 14 March 2020, at 14:43 Israel Standard Time (Greenwich Mean Time + 2 hours), and was disseminated through social media and traditional press media. As of 23 March, 18:00 Israel Standard Time, a cumulative number of 74,256 responses had been received from 69,386 adults (93.44%) and 4,870 children (6.56%) (participant characteristics, Table 1). Of these, 3,007 respondents (4.05%) reported that they were currently in isolation, of which 1,458 (48.49%) were in isolation due to a recent international travel and 1,549 (51.51%) were in isolation due to a contact with a person with COVID-19 or a person who recently returned from abroad. A new version of the questionnaire

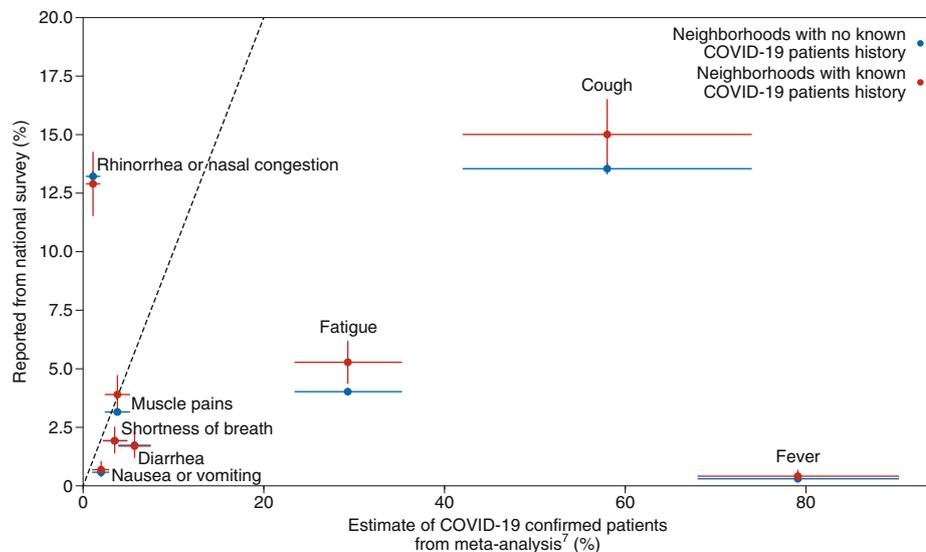


Fig. 2 | Symptom prevalence. Prevalence of symptoms for survey responses from neighborhoods in which confirmed cases were present (red) or no confirmed cases were present (blue), presented as estimates and 95% confidence intervals of patients with COVID-19 from a published meta-analysis⁸ (x axis) plotted against prevalence from survey data and bootstrap estimates of 95% confidence intervals (y axis); dashed diagonal line, $y = x$.

was established on 21 March, driven by new policies implemented by the Israeli MOH (Extended Data Fig. 1) and accumulating data on patients' symptoms⁸. The updated version includes several more questions (Extended Data Fig. 2) and has not been distributed yet.

We attempted to reach all sectors of the Israeli population in distributing the survey—first, by translating and distributing it in five languages (Hebrew, Arabic, English, Russian and Amharic) that reflect the most common languages spoken in Israel. Second, we are devoting efforts to reach underrepresented populations through several channels, including call centers, media appearance and promotion of the survey through Arabic-speaking television stations to gain interest and compliance in all sectors of the population.

We analyzed the symptoms ratio of participants by geographical location in Israel (Fig. 1). This analysis revealed differences in the proportion of reported symptoms in participants from different cities and different neighborhoods that are geographically close to each other, which might suggest the ability to detect changes at high geographical resolution.

We also analyzed the association between the prevalence of symptoms reported in the survey and the prevalence of the same symptoms in patients with COVID-19⁸. We then integrated data from the Israeli MOH on the locations of known COVID-19 cases and divided the

responses into two groups depending on whether they were living in neighborhoods in which confirmed cases were present or not. Notably, in neighborhoods in which people with confirmed COVID-19 were present, we detected a higher prevalence of symptoms that were highly prevalent in patients with confirmed COVID-19 (e.g., cough) and lower rates of symptoms that were less prevalent (e.g., rhinorrhea), which demonstrates the potential of our method for detecting disease clusters at high geographical resolution (Fig. 2).

In conclusion, we have developed a short survey based on symptoms associated with COVID-19 with the primary goal of early detection of clusters of COVID-19 outbreak. At the time of this writing, only 10 days after the survey was first distributed, 74,256 responses had been received. As expected, we also detected a higher percentage of symptoms among people who were in home isolation than among those who were not (0.06 and 0.05, respectively; $P = 5 \times 10^{-14}$ (two-sample *t*-test)).

Although the spread of COVID-19 is exponential⁹, and the number of patients with confirmed COVID-19 in Israel has increased from 193 on 14 March to 1,238 on 23 March¹⁰, it has yet to reach the vast majority of Israel's population. Thus, it is possible that our measured symptoms could be reflective of other conditions (such as influenza) that were prevalent in Israel during this period, as many diseases share common symptoms¹¹.

Our tool has several potential applications. Although it does not have the ability to diagnose individual cases of COVID-19, it might help to predict future spreading zones a few days before an outbreak occurs, with a high level of accuracy, given a sufficient sample size. Here we have provided a color map of Israel by regions of symptoms ratio (Fig. 1), but as the daily response rate increases, we expect to derive predictive models. We anticipate that these would be leveraged by policymakers to make informed decisions through the utilization of efficient regional prevention strategies rather than a uniform approach. Our survey might also be used to evaluate the effectiveness of prevention strategies implemented by public-health organizations, such as the various social-distancing measures that are currently being implemented in many countries, including Israel¹². This can be done by measuring the effect of different strategies on reducing the number of symptomatic people. Finally, it might help in elucidating the clinical course of COVID-19 by tracking the dynamics of symptoms in the population over time.

Addressing the ongoing needs of the medical and scientific community, as well as feedback from policymakers, will drive the direction and focus of our future work. To improve ease of use by participants and streamline the data-collection process, we are also building a designated mobile application that will be finalized and rolled out as soon as it is available. We also plan to resolve privacy issues around location sharing in the future application, data for which will be used only at an aggregated level and can substantially improve our models, and provide valuable insights on population interactions, adherence and disease-spread dynamics.

Our approach has many possible clinical implications; however, we have also encountered several challenges. Given that participants will be asked for personal medical information, there are concerns about identification and potential misuse of information. As mentioned above, we ask participants to fill out the survey anonymously, but we do ask for address details. These data are accessible only by the study investigators, and we are investing resources in properly securing the data to ensure that the privacy rights of our participants will be protected. Since our survey is anonymous, we cannot link the same participant's daily questionnaires, which could provide individual trends as we proceed. Another major challenge with the type of data we collect is that it is prone to selection bias. We observe that regions with relatively high response rate are regions

associated with higher socioeconomic status. Some bias may decrease as these surveys become more widely used and thus better reflect the true population; we intend to model and adjust for different factors such as age and location, and to implement national socioeconomic indices, in future analyses.

We urge other countries to adopt this tool and encourage their populations to use these daily, simple, one-minute surveys. We call for an international collaboration that will allow the sharing of methods and collected data. We also call for the large technology and social-media companies already collecting elements of personal data to collaborate in this international effort by sharing regional information to help us improve our models.

Ethics declaration

The study protocol was approved by the institutional review board (IRB). Informed consent was waived by the IRB, as all identifying details of the participants were removed before the computational analysis. Participants were made fully aware of the way in which the data will be stored, handled and shared, which was provided to them and is in accord with the privacy and data-protection policy of the Weizmann Institute of Science (<https://weizmann.ac.il/pages/privacy-policy>).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Tables of de-identified, aggregated data are available at <https://github.com/hrossman/Covid19-Survey>. Map data are copyrighted by

OpenStreetMap contributors and are available from <https://www.openstreetmap.org>.

Code availability

Analysis source code is available at <https://github.com/hrossman/Covid19-Survey>. Source code for the questionnaires is available at <https://github.com/hasadna/avid-covider> as an open source project, and can be readily adapted to use in other countries. □

Hagai Rossman^{1,2}, Ayya Keshet^{1,2}, Smadar Shilo^{1,2,3}, Amir Gavrieli^{1,2}, Tal Bauman⁴, Ori Cohen^{1,2}, Esti Shelly⁵, Ran Balicer⁶, Benjamin Geiger², Yuval Dor⁷ and Eran Segal^{1,2} ✉

¹Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. ²Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. ³Pediatric Diabetes Unit, Ruth Rappaport Children's Hospital, Rambam Healthcare Campus, Haifa, Israel. ⁴Mapping and Geo-Information Engineering, Civil and Environmental Engineering Faculty, The Technion, Israel. ⁵Ministry of Health, Jerusalem, Israel. ⁶Clalit Research Institute, Clalit Health Services, Tel Aviv, Israel. ⁷School of Medicine-IMRIC-Developmental Biology and Cancer Research, The Hebrew University, Jerusalem, Israel. ✉e-mail: eran.segal@weizmann.ac.il

Published online: 09 April 2020
<https://doi.org/10.1038/s41591-020-0857-9>

References

- Guan, W.-J. et al. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2002032> (2020).
- World Health Organization. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (accessed 24 March 2020).
- Klompas, M. *Ann. Intern. Med.* <https://doi.org/10.7326/M20-0751> (2020).
- Davis, N. *The Guardian* <https://www.theguardian.com/science/2020/mar/24/uk-app-aims-to-help-researchers-track-spread-of-coronavirus> (2020).

- Segal, E. et al. *medRxiv* <https://doi.org/10.1101/2020.04.02.20051284> (2020).
- State of Israel Ministry of Health. <https://www.health.gov.il/English/Topics/Diseases/corona/Pages/default.aspx> (accessed 24 March 2020).
- Corman, V. M. et al. *Euro Surveill.* **25**, 2000045 (2020).
- Zhao, X. et al. *medRxiv* <https://doi.org/10.1101/2020.03.17.20037572> (2020).
- Li, Y. et al. *medRxiv* <https://doi.org/10.1101/2020.03.01.20029819> (2020).
- State of Israel Ministry of Health. <https://www.health.gov.il/English/Topics/Diseases/corona/Pages/press-release.aspx> (accessed 24 March 2020).
- Zhang, H. et al. *Preprints* <https://doi.org/10.20944/preprints202003.0160.v1> (2020).
- Buckee, C. O. et al. *Science* <https://doi.org/10.1126/science.abb8021> (2020).

Acknowledgements

We thank U. Feinstein for assisting us in defining the surveys, symptoms and medical conditions. We thank F. Zhang, O. Shalem, W. Allen, B. Silbermann, R. Probasco and D. Cheng for insightful discussions and look forward to jointly creating an international consortium with them. We thank the following for their contributions to our efforts: T. Meir, I. Kalka, A. Lavon, T. Karady, A. Godneva, D. Kolobkov, S. Shoer, O. Bartal and the people at Israel Corona Map; and T. Ben-Ami, M. Hashkes, H. Ben-Shushan, R. Miara, T. Eldar, S. Kasem, T. Bria, S. Avraham, B. Kirel, A. Terkeltaub, D. Hizi, A. Kariv, M. Zer-Aviv, N. Kastel, R. Folkman, G. Barabash and the Public Knowledge Workshop ('Hasadna').

Author contributions

H.R., A.K., S.S. and A.G. conceived of the project, designed and conducted the analyses, interpreted the results and wrote the manuscript; T.B. designed and conducted the analyses and interpreted the results; O.C. directed the organizational and logistic efforts and interpreted the results; E.S. provided and interpreted data from the MOH; and R.B., B.G., Y.D. and E.S. conceived of and directed the project and analyses, designed the analyses, interpreted the results, wrote the manuscript and supervised the project.

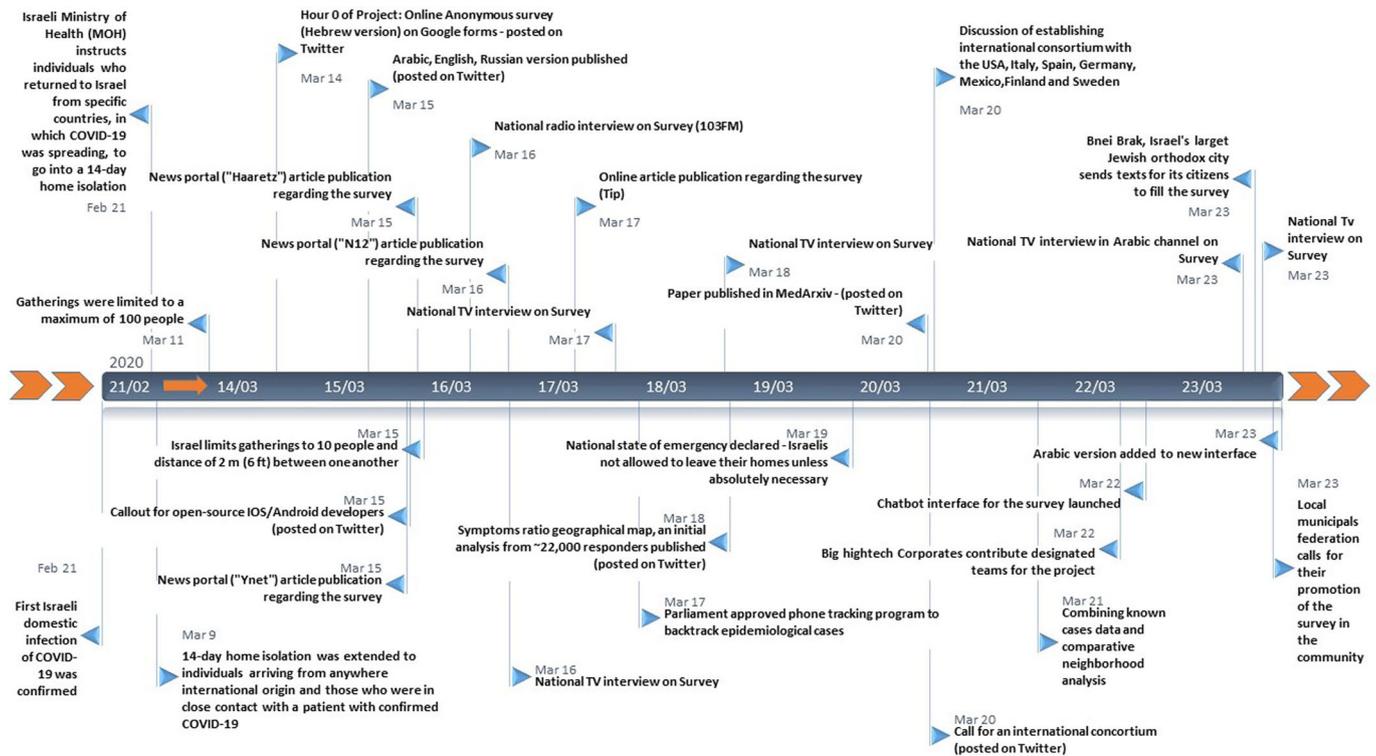
Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41591-020-0857-9>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-0857-9>.



Extended Data Fig. 1 | Project timeline describing all major events in its development including national events which affected and drove its process, from time survey online publication (March 14th, 14:44) to March 23rd, March 21, 18:00 IST.

COVID-19 Daily questionnaire

This questionnaire was designed to detect potential geographic areas in which the coronavirus is spreading in Israel. You will be asked about potential symptoms of the virus. A better estimation of the number of infected people in each area will help to identify locations in which the number of infected people is particularly high. Please fill the questionnaire every day for each family member separately. Please fill it also in cases you and your family are feeling well and do not experience any symptoms. Of note, this questionnaire can not diagnose a coronavirus infection. It is anonymized and all the data will be used solely for epidemiologic purposes. We are taking every measure to keep the privacy of the responders.

We thank you for your participation

&*Age

&*Gender:

- Male
 Female

&*City, Street

#*I am:

- Feeling well
 Not feeling well

*Are you experiencing any of the following symptoms?

- Cough:
 #Dry cough (no sputum)
 #Wet cough (with sputum)
- Fatigue
 Muscle pain
 Shortness of breath
 Rhinorrhea (Runny nose) and/ or Nasal congestion
 Diarrhea
 Nausea and /or vomiting
 #Sore throat
 #Headache
 #Chills
 #Confusion
 #Experiencing loss of taste and/or smell sensation

&*Have you been diagnosed with any of the following conditions:

- Diabetes mellitus
 Hypertension
 Cardiovascular disease or stroke
 Chronic lung disease including Asthma (with the exception of childhood Asthma)
 Chronic kidney disease
 #Malignancy (cancer)
 #Immunodeficiency (including consumption of drugs which cause immunodeficiency)

*I am currently:

- Not in isolation
 In isolation (including from family members, staying in a separate room) from the date of _____ - due to :
 A recent international travel
 A contact with an individual who was infected with coronavirus or an individual who recently returned from any destination abroad
 #Experiencing disease symptoms
 #Voluntary isolation
- #I have a confirmed infection with COVID-19 (by a lab test) and currently:
 In home isolation
 Staying in a hotel
 Hospitalized in a hospital
 I recovered from COVID-19 infection and staying at home

&*Cigarette smoking habits

- I currently smoke
 I used to smoke and stopped more than 5 years ago
 I used to smoke and stopped less than 5 years ago
 I have never smoked

What is your current body temperature?

- I did not measure my temperature in the last 24 hours
 I measured my temperature and it was _____

#*With how many individuals have you been in contact in the last 24 hours? (within approximately 2 meters (6 ft 7 in) for more than 15 minutes)

Adults (age above 18 years old _____)

Children (age below 18 years old _____)

Extended Data Fig. 2 | COVID-19 Daily questionnaire. The most updated version is presented #Questions that were added in the new version of the questionnaire and are therefore not analyzed in this paper *Questions that the participant is required to answer, &Questions that should be filled only once.