medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

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

Hagai Rossman[1,2], Ayya Keshet[1,2], Smadar Shilo[1,2,3], Amir Gavrieli[1,2], Tal Bauman[4], Ori Cohen, Ran Balicer[5], Benjamin Geiger[2], Yuval Dor[6], Eran Segal[1,2,†]

[1]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel

[2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel

[3]Pediatric Diabetes Unit, Ruth Rappaport Children's Hospital, Rambam Healthcare Campus, Haifa, Israel

[4]Mapping and Geo-Information Engineering, Civil and Environmental Engineering Faculty, The Technion, Israel

[5]Clalit Research Institute, Clalit Health Services, Tel Aviv, Israel

[6]School of Medicine-IMRIC-Developmental Biology and Cancer Research, The Hebrew University

Abstract

Coronavirus infection spreads in clusters and therefore 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 to inform epidemiologists, public health officials, and policy makers. We show preliminary results from a survey of over 38,000 Israelis and call for an international consortium to extend this concept in order to develop predictive models. We expect such data to allow: Faster detection of spreading zones and patients; Obtaining a current snapshot of the number of people in each area who have developed symptoms; Predicting future spreading zones several days before an outbreak occurs; Evaluating the effectiveness of the various social distancing measures taken, and their contribution to reduce the number of symptomatic people. Such information can provide a valuable tool for decision makers to decide which areas need strengthening of social distancing measures and which areas can be relieved. Researchers from the U.S, Spain, and Italy have adopted our approach and we are collaborating to further improve it. We call with urgency for other countries to join this international consortium, and to share methods and data collected from these daily, simple, one-minute surveys.

Introduction

In December 2019 a novel coronavirus was isolated, after a cluster of patients with pneumonia of unknown cause were diagnosed in China. This new isolate was named 2019-nCoV and is the cause for the COVID-19 disease (Na Zhu et al. 2019). The virus has led to an ongoing outbreak and an international health crisis. Update reports of the World Health Organization (WHO) state that at the moment there are at the low hundreds of thousands confirmed cases and thousands of deaths of patients who were infected by COVID-19 globally [1]. This number is rapidly increasing and most probably is a vast underestimation of the real number of patients worldwide, as infected individuals are contagious even when minimally symptomatic or asymptomatic [2]. The spread of the disease has presented an extreme challenge to the international community. Although the WHO issues health policies recommendations, many policy-makers from different countries have chosen and implied different strategies. These strategies depend on many factors, including the local spread of COVID-19, healthcare system resources, economical and political factors, public adherence and their perception of the situation.

One of the major challenges to date is disease detection and diagnosis. While the gold standard for COVID-19 diagnosis is detecting the virus by a real-time RT-PCR testing [3], current resource and policy

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

limitations in many countries restrict the amount of testing that can be performed . As testing the entire population for the presence of coronavirus is not feasible, here we have developed a simple one-minute online questionnaire with the goal of early and temporal detection of geographic clusters in which the virus is spreading. This additional information provides real time analysis of symptoms over the course of days as people are getting sick and render an informative situation snapshot.

Methods

We developed an online daily survey (http://predict-corona.org/) designed to locate potential geographic areas in which the coronavirus is spreading throughout Israel. The survey contains questions on age, gender, geographic location (city, street, zip code), isolation status and smoking habits. Furthermore, responders were asked to report whether they experience symptoms which were defined as common symptoms of the Coronavirus by healthcare professionals based on the existing literature [4,5]. These include cough, fatigue, myalgia (muscle pain), shortness of breath, rhinorrhea or nasal congestion, diarrhea and nausea or vomiting. In addition, participants were asked about the existence of one of the following chronic health conditions: Diabetes mellitus, Hypertension, Ischemic heart disease, Asthma, chronic lung disease and chronic kidney disease. Participants were also asked to measure their daily body temperature and document it in the questionnaire (Figure 1). The survey is distributed in 5 languages - Hebrew, Arabic, English, Russian, and Amharic to reflect and serve the diverse spoken languages of Israel's population.

For each responder we calculate the *symptoms ratio*:

$$\textit{Symptoms ratio} = (\#\ \textit{reported symptoms})/6\ \textit{if age} <= 18\ \textit{else}\ (\#\ \textit{reported symptoms})/5$$

The number of reported symptoms divided by the number of symptoms in our predefined list. Symptoms in this list were predefined by the Israeli MOH (Ministry of Health): muscle pains, shortness of breath, fatigue, cough and a high fever (body temperature of over 38 degrees celsius. For responders under the age of 18 nausea and vomiting was also included.

Responders were associated with an area in Israel using the address provided to create a colormap of Israel by the aggregated symptoms ratio as defined by the MOH.

Figure 1: Daily questionnaire

This questionnaire was designed to locate 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,

Prof Eran Segal, Weizmann Institute of Science
Prof Yuval Dor, The Hebrew University of Jerusalem
Prof Benjamin Geiger, Weizmann Institute of Science

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Age

Gender:

- ❏ Male
- ❏ Female

City, Street, Zip Code

Are you experiencing any of the following symptoms?

- ❏ I'm feeling well
- ❏ Cough
- ❏ Fatigue
- ❏ Muscle pain
- ❏ Shortness of breath
- ❏ Rhinorrhea (Runny nose) or Nasal congestion
- ❏ Diarrea
- ❏ Nausea and vomiting

Have you been diagnosed with any of the following conditions:
Diabetes mellitus, Hypertension, Ischemic heart disease, Asthma, chronic lung disease, Chronic kidney disease

- ❏ Yes
- ❏ No

I am currently:

- ❏ Not in isolation
- ❏ In isolation - due to a recent international travel
- ❏ In isolation - due to a contact with an individual who was infected with coronavirus or an individual who recently returned from any destination abroad

Cigarette smoking habits

- ❏ I currently smoke
- ❏ I used to smoke
- ❏ I have never smoked

What is your current body temperature? (Celsius)

Results

The questionnaire was first distributed on March 14th 2020 at 14:43 (Israeli time zone, GMT+2) and was spread via social media and press. Of note, the first case of confirmed coronavirus in Israel was confirmed on February 21st 2020. To this date, March 19th, 12pm, there have been 38,320 responses. The characteristics of the responders are described in Table 1. The Israeli MOH guidelines have instructed individuals who return to Israel from anywhere in the world to be placed under home isolation for 14 days from their date of return to Israel starting from March 09, 2020. In addition, individuals that were in close contact as defined by being within approximately 2 meters (6 feet) of a COVID-19 case for more than 15 minutes were also instructed to be in home isolation for a similar time period [6]. Altogether, 1,509 (4.02%) responders reported that they are currently in isolation. For those reported as being in isolation, 774 (51.3%) responders were due to a recent international travel and 751 (49.7%) were due to a contact with an individual who was infected with coronavirus or an individual who recently returned from any destination abroad.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Table 1. Baseline characteristics of questionnaire responders

Characteristic, mean (SD) or %	All (n =38,320)	Not in home isolation (n = 36,811) (95.98%)	In home isolation (n = 1,509) (4.02%)	2020-03-14 (n = 884) (2.31%)	2020-03-15 (n = 4,399) (11.48%)	2020-03-16 (n = 9,427) (24.6%)	2020-03-17 (n = 8,617) (22.49%)	2020-03-18 (n = 13,564) (35.4%)	2020-03-19 (n = 1,429) (3.73%)
Age (years)	43.74 (17.71)	43.88 (17.72)	40.33 (17.06)	39.49 (17.77)	40.07 (15.45)	41.95 (16.45)	46.93 (18.68)	44.22 (18.15)	45.64 (17.96)
Sex - Male	18,131 (47.31%)	17,366 (47.18%)	765 (50.7%)	404 (45.7%)	2,415 (54.9%)	4,456 (47.27%)	3,920 (45.49%)	6,272 (46.24%)	664 (46.47%)
Smoking history (previously smoked/ currently smoking)	13,838 (36.11%)	13,336 (36.23%)	502 (33.27%)	305 (34.5%)	1,461 (33.21%)	3,413 (36.2%)	3,113 (36.13%)	4,995 (36.83%)	551 (38.56%)
Presence of a chronic medical conditions	7,106 (18.54%)	6,891 (18.72%)	215 (14.25%)	139 (15.72%)	649 (14.75%)	1,543 (16.37%)	1,846 (21.42%)	2,647 (19.51%)	282 (19.73%)
Symptoms									
No symptoms (Feel good)	31,435 (82.03%)	30,243 (82.16%)	1,192 (78.99%)	696 (78.73%)	3,501 (79.59%)	7,633 (80.97%)	7,212 (83.7%)	11,236 (82.84%)	1,157 (80.97%)
Body temperature (Celcius)	36.52 (0.45)	36.52 (0.45)	36.56 (0.48)	36.56 (0.44)	36.55 (0.43)	36.55 (0.46)	36.52 (0.44)	36.49 (0.45)	36.45 (0.49)
Body temperature equals or above 38 degrees celsius	142 (0.37%)	123 (0.33%)	19 (1.26%)	2 (0.23%)	17 (0.39%)	47 (0.5%)	24 (0.28%)	44 (0.32%)	8 (0.56%)
Nausea and vomiting	240 (0.63%)	227 (0.62%)	13 (0.86%)	8 (0.9%)	32 (0.73%)	65 (0.69%)	47 (0.55%)	77 (0.57%)	11 (0.77%)
Muscle pain	1,324 (3.46%)	1,247 (3.39%)	77 (5.1%)	29 (3.28%)	166 (3.77%)	365 (3.87%)	267 (3.1%)	444 (3.27%)	53 (3.71%)
Rhinorrhea or nasal congestion	5,386 (14.06%)	5,177 (14.06%)	209 (13.85%)	157 (17.76%)	703 (15.98%)	1,372 (14.55%)	1,167 (13.54%)	1,768 (13.03%)	219 (15.33%)
Fatigue	1,667 (4.35%)	1,573 (4.27%)	94 (6.23%)	57 (6.45%)	220 (5.0%)	477 (5.06%)	333 (3.86%)	521 (3.84%)	59 (4.13%)
Shortness of breath	791 (2.06%)	741 (2.01%)	50 (3.31%)	18 (2.04%)	102 (2.32%)	225 (2.39%)	170 (1.97%)	250 (1.84%)	26 (1.82%)
Cough	5,610 (14.64%)	5,351 (14.54%)	259 (17.16%)	144 (16.29%)	667 (15.16%)	1,442 (15.3%)	1,170 (13.58%)	1,975 (14.56%)	212 (14.84%)
Diarrhea	706 (1.84%)	660 (1.79%)	46 (3.05%)	24 (2.71%)	97 (2.21%)	189 (2.0%)	144 (1.67%)	227 (1.67%)	25 (1.75%)
Symptoms ratio									
Symptoms ratio	0.05 (0.11)	0.05 (0.11)	0.07 (0.15)	0.06 (0.13)	0.05 (0.12)	0.05 (0.12)	0.05 (0.11)	0.05 (0.11)	0.05 (0.12)

* Symptoms ratio was calculated as the number of reported symptoms divided by the number of symptoms in the predefined MOH symptoms list.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

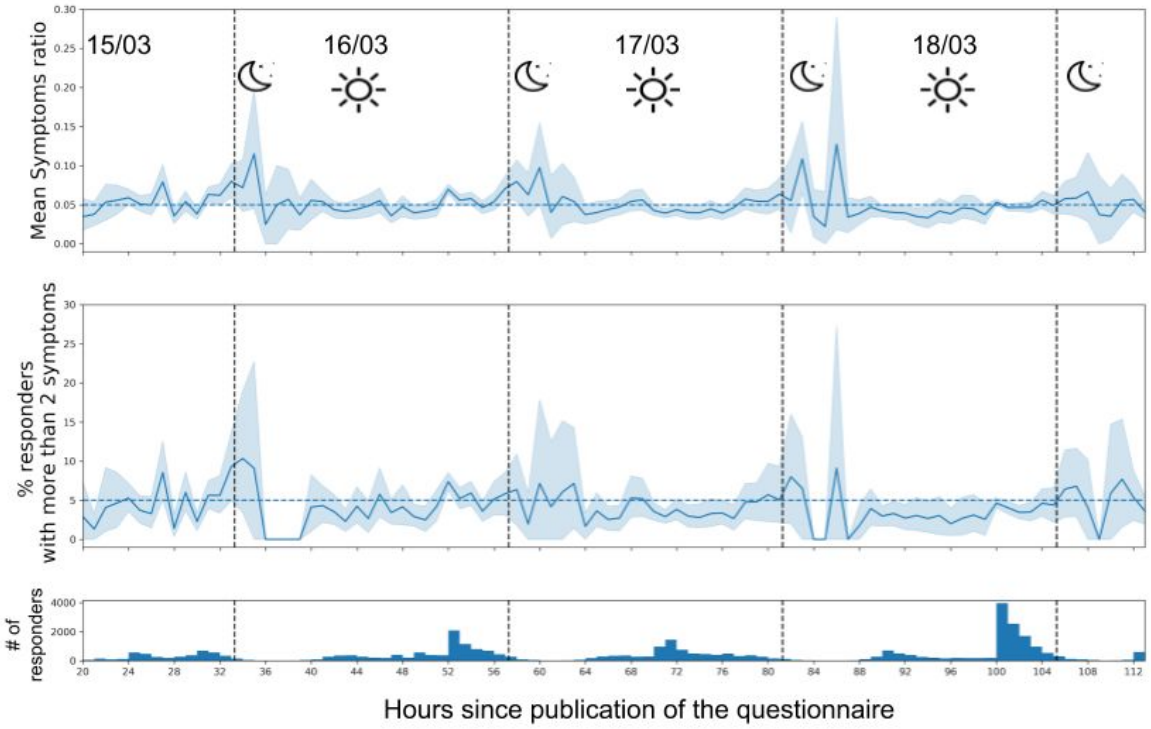

Figure 2: Nationwide percentage of symptoms since publishing questionnaire. **Top**: Mean symptoms ratio of responders. **Middle**: Percentage of responders who reported more than 2 symptoms. **Bottom**: Number of responding participants per hour.

We next analysed the symptoms ratio of responders by geographical locations in Israel (Figure 3).

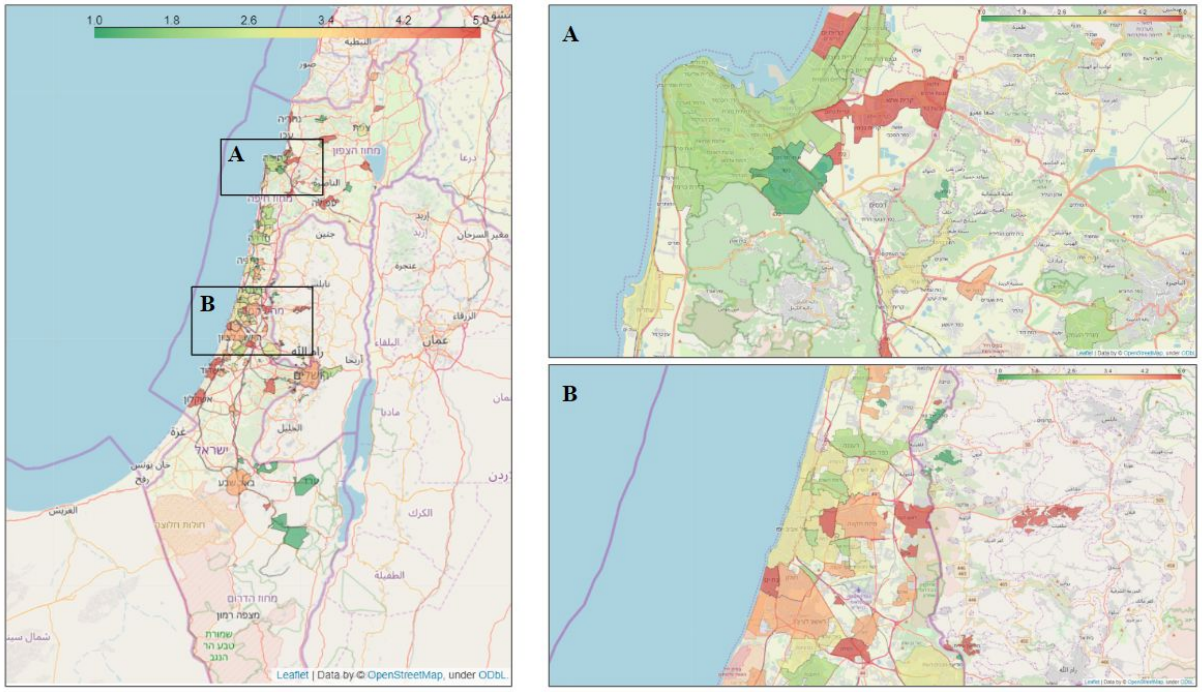

Figure 3: Average COVID-19 associated symptoms region map. City municipal regions with at least 30 responders are shown. Each region is colored by a category defined by the average symptoms ratio, calculated by averaging the reported symptoms rate by responders in that city. Green - low symptoms rate, red - high symptoms rate.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Discussion

Here, we developed a short survey based on symptoms associated with COVID-19 infection with the primary goal of early detection of clusters of COVID-19 infection. At the time of writing, only five days after the survey has been distributed, 38,320 responders have completed it. Our initial analysis based on the preliminary data already reveals a different rate of symptoms in responders from different cities which are geographically close to each other, which might hint at the ability to detect changes in a high geographical resolution (figure 3). As expected, we are also detecting a higher percentage of symptoms among individuals who are in home isolation, compared to those who are not (0.07 and 0.05 respectively).

Temporal analysis of the symptoms rate in these first 5 days shows a relatively stable rate in the population at daytime, where response rate is relatively high (Figure 2). Although the spread of COVID-19 infection is exponential [7], and the number of confirmed COVID-19 patients in Israel has increased from 195 cases on March 14th 2020 to 529 patients on March 19th noontime [8], it has yet to reach the vast majority of Israel's population. We therefore hypothesize that these symptoms may reflect other respiratory infections which are prevalent in Israel during this period (such as an infection caused by an influenza virus), as many of these diseases share common symptoms [9]. If our hypothesis is correct, this rate can be viewed as a baseline rate that will allow us to further detect subtle changes which may reflect a spread of COVID-19 infection. These findings encourage us to reach as much of the population as we can in a minimal amount of time to actively create a reliable baseline of the symptoms for the population.

Our tool has several potential applications. First, while it does not have the ability to diagnose individual cases of COVID-19 infection, it may help predict future spreading zones a few days before an outbreak occurs, with a high level of accuracy given a sufficient sample size. We currently provide a colormap of Israel by regions of symptoms ratio (figure 3), however, as the daily response rate will increase, enough data will be collected, enabling us to implement prediction models. These we hope would be leveraged by policy-makers to make informed decisions, by utilization of efficient regional prevention strategies rather than a universal approach. Second, it may also be used to evaluate the effectiveness of prevention strategies implemented by public health organizations, such as the various social distancing measures which are currently being employed in many countries, including Israel. This can be done by measuring the effect of different strategies on reducing the number of symptomatic individuals. Third, it may help in understanding the clinical course of COVID-19 infection 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 of policy-makers will drive the directions and the focus of our future work. To improve ease of use by responders and streamline the data collection process, we are also building a designated mobile application which is being finalized as these lines are written. We also hope to resolve privacy issues around location sharing in the future application, which will be used ONLY in an aggregated level and can significantly improve our models, and provide valuable insights on population interactions, adherence and disease spread dynamics.

Although our approach has many possible clinical implications, we have so far encountered several challenges. Asking participants to fill the survey may raise issues of information privacy related to the collection of medical data, to avoid any privacy issues that may occur, our survey is filled anonymously. Moreover, since our survey is anonymous we cannot link the same responder's daily questionnaires, which can provide an individual trend as we proceed. However, our goal is to provide a macro-management tool which is still feasible. Finally, a main challenge when relying on this data is that it is baised. We notice that regions with

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

relatively high response rate are regions associated with higher socioeconomic status. It may also be assumed that people who answer these surveys tend to be more adherent to healthcare protocols and thus less likely to be infected. While some of the bias may decrease as these surveys become more popular and known and thus better reflect the true population, we intend to adjust for different factors such as age and location.

In summary, we present a new tool that has the potential to early detect clusters of COVID-19 infection. This paper and the presented analyses were written in great haste and urgency. We were inspired by the WHO executive director Dr. Michael Ryan words: "*Perfection is the enemy of the good when it comes to emergency management*". We urge other countries to adopt this tool and encourage their population to use these daily, simple, one-minute surveys. We call for an international collaboration, which will allow the sharing of methods and collected data. We also call for the "Tech Giants" Google, Facebook and Twitter to collaborate in this international effort by sharing REGIONAL (not personal) information to help us improve our models.

Acknowledgments

We thank Dr. Uri Feinstein for assisting us in defining the surveys, symptoms, and medical conditions. We thank Feng Zhang, Ophir Shalem, and Will Allen for insightful discussions and look forward to jointly creating an international consortium with them.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

References

1. www.who.int/docs/default-source/coronaviruse/situation-reports/20200317-sitrep-57-covid-19.pdf.
2. Klompas, M. Coronavirus Disease 2019 (COVID-19): Protecting Hospitals From the Invisible. *Ann. Intern. Med.* (2020). doi:10.7326/M20-0751
3. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* **25,** (2020).
4. Guan, W.-J. *et al.* Clinical characteristics of coronavirus disease 2019 in china. *N. Engl. J. Med.* (2020). doi:10.1056/NEJMoa2002032
5. Norris, J. M., Johnson, R. K. & Stene, L. C. Type 1 diabetes-early life origins and changing epidemiology. *Lancet Diabetes Endocrinol.* (2020). doi:10.1016/S2213-8587(19)30412-7
6. The Novel Coronavirus, Ministry of Health. at <https://www.health.gov.il/English/Topics/Diseases/corona/Pages/default.aspx>
7. Li, Y. *et al.* COVID-19 Epidemic Outside China: 34 Founders and Exponential Growth. *medRxiv* (2020). doi:10.1101/2020.03.01.20029819
8. COVID-19 Press Releases, Ministry of Health. at <https://www.health.gov.il/English/Topics/Diseases/corona/Pages/press-release.aspx>
9. Zhang, H. *et al.* Influenza-Like Illness Caused by the 2019 Novel Coronavirus (2019-nCoV) via the Person-to-Person Transmission. (2020). doi:10.20944/preprints202003.0160.v1

medRxiv preprint doi: https://doi.org/10.1101/2020.03.19.20038844; this version posted March 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license .

Supplementary material

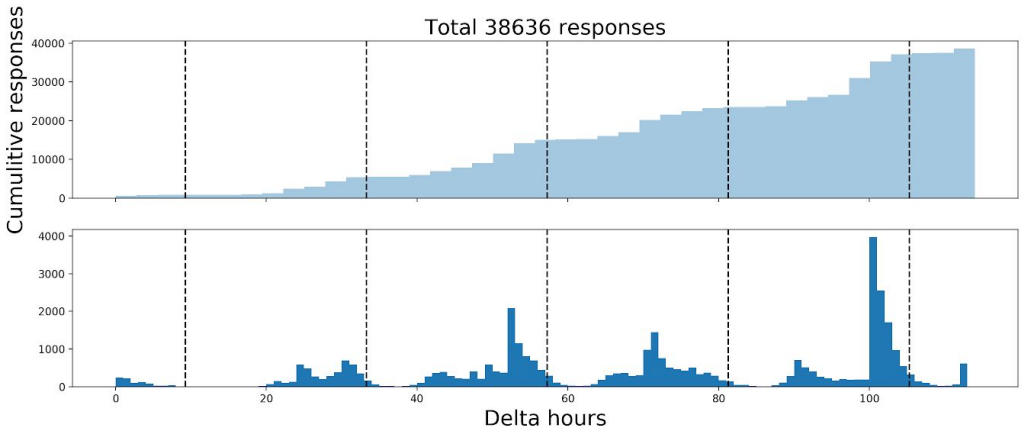

Figure 4: Number of responding participants per hour. **Top**: Cumulative. **Bottom**: Histogram.

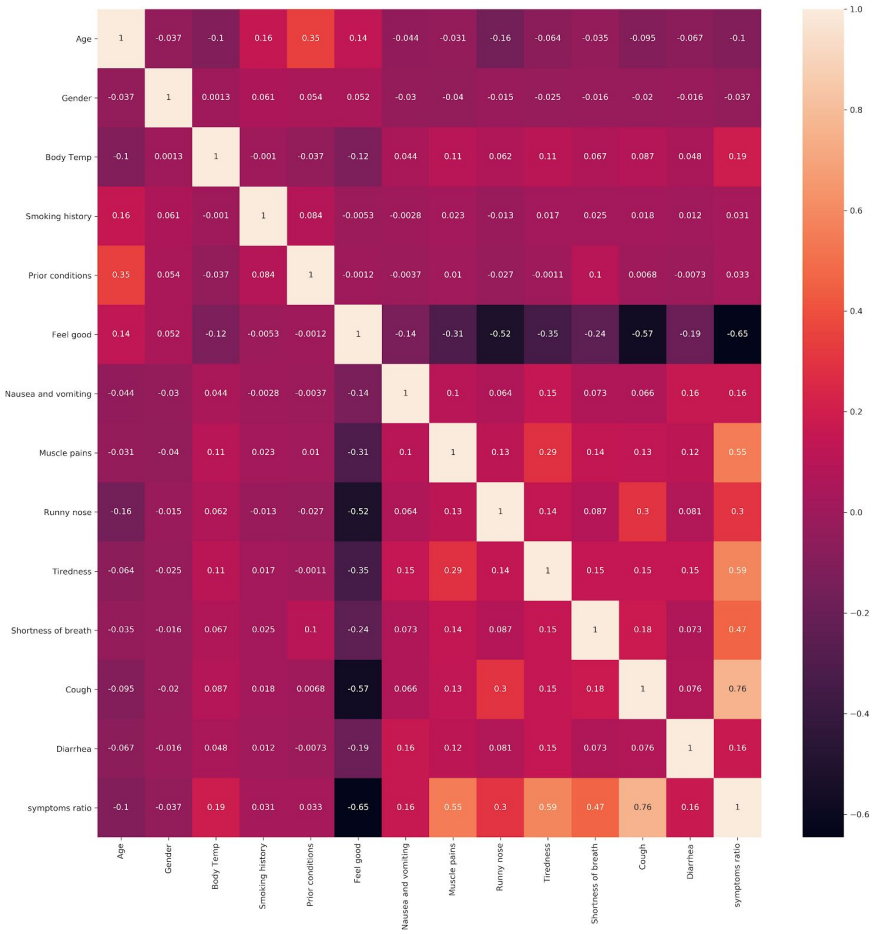

Figure 5: Correlation heatmap of responder characteristics and symptoms