



Wearable sensor data and self-reported symptoms for COVID-19 detection

Giorgio Quer^{1,3}  , Jennifer M. Radin^{1,3} , Matteo Gadaleta^{1,3} , Katie Baca-Motes¹, Lauren Ariniello¹ , Edward Ramos^{1,2}, Vik Kheterpal^{1,2} , Eric J. Topol¹  and Steven R. Steinhubl¹ 

Traditional screening for COVID-19 typically includes survey questions about symptoms and travel history, as well as temperature measurements. Here, we explore whether personal sensor data collected over time may help identify subtle changes indicating an infection, such as in patients with COVID-19. We have developed a smartphone app that collects smartwatch and activity tracker data, as well as self-reported symptoms and diagnostic testing results, from individuals in the United States, and have assessed whether symptom and sensor data can differentiate COVID-19 positive versus negative cases in symptomatic individuals. We enrolled 30,529 participants between 25 March and 7 June 2020, of whom 3,811 reported symptoms. Of these symptomatic individuals, 54 reported testing positive and 279 negative for COVID-19. We found that a combination of symptom and sensor data resulted in an area under the curve (AUC) of 0.80 (interquartile range (IQR): 0.73–0.86) for discriminating between symptomatic individuals who were positive or negative for COVID-19, a performance that is significantly better ($P < 0.01$) than a model¹ that considers symptoms alone (AUC = 0.71; IQR: 0.63–0.79). Such continuous, passively captured data may be complementary to virus testing, which is generally a one-off or infrequent sampling assay.

Owing to the current lack of fast and reliable testing, one of the greatest challenges for preventing transmission of SARS-CoV-2 is the ability to quickly identify, trace and isolate cases before they can further spread the infection to susceptible individuals. As regions across the United States start implementing measures to reopen businesses, schools and other activities, many rely on current screening practices for COVID-19, which typically include a combination of symptom and travel-related survey questions and temperature measurements. However, this method is likely to miss pre-symptomatic or asymptomatic cases, which make up ~40–45% of those infected with SARS-CoV-2, and who can still be infectious^{1,2}. An elevated temperature ($>100^{\circ}\text{F}$ ($>37.8^{\circ}\text{C}$)) is not as common as frequently believed, being present in only 12% of individuals who tested positive for COVID-19³ and just 31% of patients hospitalized with COVID-19 (at the time of admission)⁴.

Smartwatches and activity trackers, which are now worn by one in five Americans⁵, can improve our ability to objectively characterize each individual's unique baseline for resting heart rate⁶, sleep⁷ and activity and can therefore be used to identify subtle changes in that user's data that may indicate that they are coming down with a viral illness. Previous research from our group has shown that this method, when aggregated at the population level, can

significantly improve real-time predictions for influenza-like illness⁸. Consequently, we created a prospective app-based research platform, called DETECT (Digital Engagement and Tracking for Early Control and Treatment), where individuals can share their sensor data, self-reported symptoms, diagnoses and electronic health record data with the aim of improving our ability to identify and track individual- and population-level viral illnesses, including COVID-19.

A previously reported study that captured symptom data in over 18,000 SARS-CoV-2-tested individuals via a smartphone-based app found that symptoms were able to help distinguish between individuals with and without COVID-19¹. The aim of this study is to investigate if the addition of individual changes in sensor data to symptom data can be used to improve our ability to identify COVID-19-positive versus COVID-19-negative cases among participants who self-reported symptoms.

Between 25 March and 7 June 2020, our research study enrolled 30,529 individuals, with representation from every state in the United States. Among the consented individuals, 62.0% are female and 12.8% are 65 or more years old. Of the participants, 78.4% connected their Fitbit devices to the study app, 31.2% connected the data from the Apple HealthKit, while 8.1% connected data from Google Fit (note that an individual can connect to multiple platforms). In addition, 3,811 reported at least one symptom (12.5%); of those, 54 also reported testing positive for COVID-19 and 279 reported testing negative. The numbers of days per different data type and data aggregator system are reported in Table 1, while the symptoms distribution for symptomatic individuals tested for COVID-19, or not tested, is shown in Fig. 1.

A minority of symptomatic participants (30.3%) who tested for COVID-19 had a resting heart rate (RHR) greater than two standard deviations above the average baseline value during symptoms. The change in RHR on its own (Table 1) did not allow significant discrimination between COVID-19-positive and COVID-19-negative participants using the RHRMetric (area under the curve (AUC) of 0.52 (interquartile range (IQR): 0.41–0.64)) (Fig. 2a).

Sleep and activity did show a significant difference among the two groups (Table 1), with an AUC of 0.68 (0.57–0.79) for the SleepMetric (Fig. 2b) and 0.69 (0.61–0.77) for the ActivityMetric (Fig. 2c), supporting that the sleep and activity of COVID-19 positive participants were impacted significantly more than COVID-19-negative participants. Sleep and activity are slightly correlated, with a negative correlation coefficient of -0.28 , $P < 0.01$.

To evaluate the contribution of all the data types commonly available through personal devices, we combined the RHR, sleep and activity metrics in a single metric (SensorMetric, Fig. 2d). This

¹Scripps Research Translational Institute, La Jolla, CA, USA. ²CareEvolution, Ann Arbor, MI, USA. ³These authors contributed equally: Giorgio Quer, Jennifer M. Radin, Matteo Gadaleta. ✉e-mail: gquer@scripps.edu

Table 1 | Participants' characteristics and device data

	Total	Any symptom	COVID-19 positive	COVID-19 negative	P value
Demographic data					
Number of participants	30,529	3,811	54	279	–
Female	18,922 (62.0%)	2,828 (74.2%)	43 (79.6%)	199 (71.3%)	0.65
Age group (years)					
Under 35	7,052 (23.1%)	1,184 (31.1%)	17 (31.5%)	71 (25.4%)	0.52
35 to 50	10,357 (33.9%)	1,522 (39.9%)	23 (42.6%)	121 (43.4%)	1.00
51 to 65	9,038 (29.6%)	857 (22.5%)	12 (22.2%)	64 (22.9%)	1.00
Over 65	3,899 (12.8%)	219 (5.7%)	1 (1.9%)	21 (7.5%)	0.22
Fitbit users	23,922 (78.4%)	3,306 (86.7%)	46 (85.2%)	237 (84.9%)	1.00
Apple users	9,522 (31.2%)	1,219 (32.0%)	21 (38.9%)	94 (33.7%)	0.66
Sensor data					
Available days (IQR)					
RHR	322 (131–387)	325 (153–396)	312 (98–377)	300 (119–392)	0.70
Sleep	249 (48–373)	273 (105–383)	283 (52–362)	246 (70–375)	0.56
Activity	394 (370–412)	407 (379–415)	404 (375–410)	401 (374–413)	0.57
Mean change (s.d.)					
RHR (b.p.m.)	0.0 (2.84)	0.40 (3.18)	1.15 (4.83)	0.61 (3.68)	0.33
Sleep (min)	0 (54)	3 (59)	57 (92)	4 (68)	<0.01
Activity (steps)	52 (2,659)	–323 (2,771)	–3,533 (4,418)	–208 (3,086)	<0.01

Summary of the collected data and demographic information about the cohort. Available days are specified for each data type, with median and interquartile range (IQR) values. Mean changes (and s.d.) in resting heart rate (RHR), steps and sleep from baseline (–21 to –7 days) to symptomatic period (0 to 7 days) are reported; for individuals with no symptoms we consider 6 March 2020 as day 0. P values of a two-sided Fisher's exact test applied to COVID-19-positive and COVID-19-negative participants are also reported.

improved the overall performance from the three sensor metrics to an AUC of 0.72 (0.64–0.80).

We also considered a model based only on self-reported symptoms (SymptomMetric, Fig. 2e), along with age and sex. With respect to the previously published model¹, we measure a slightly lower AUC of 0.71 (0.63–0.79).

When participant-reported symptoms and sensor metrics are jointly considered in the analysis (OverallMetric, Fig. 2f), the achieved performance was significantly improved ($P < 0.01$) relative to either alone, with an AUC of 0.80 (0.73–0.86).

Discussion

Our results show that individual changes in physiological measures captured by most smartwatches and activity trackers are able to significantly improve the distinction between symptomatic individuals with and without a diagnosis of COVID-19 beyond symptoms alone. Although encouraging, these results are based on a relatively small sample of participants.

This work builds on our earlier retrospective analysis demonstrating the potential for consumer sensors to identify individuals with influenza-like illness, which has subsequently been replicated in a similar analysis of over 1.3 million wearable users in China for predicting COVID-19^{8,9}. In response to the COVID-19 pandemic, a number of prospective studies, led by device manufacturers and/or academic institutions, including DETECT, have accelerated deployment to allow interested individuals to voluntarily share their sensor and clinical data to help address the global crisis^{10–14}. The largest of these efforts, Corona-Datenspende, was developed by the Robert Koch Institut in Germany and has enrolled over 500,000 volunteers¹⁵.

As different individuals experience a wide range of symptomatic and biological responses to infection with SARS-CoV-2, it is likely that their measurable physiological changes will also vary^{16–18}. For that reason, it is possible that biometric changes may be more

valuable in identifying those at highest risk for decompensation rather than just a dichotomous distinction in infection status. Because of the limited testing in the United States, especially early in the spread of the COVID-19 pandemic, individuals with more severe symptoms may have been more likely to be tested. In fact, the majority of symptomatic participants in our study did not undergo testing. However, using the optimal tradeoff of sensitivity and specificity on the ROC, we would predict that, of the 3,478 symptomatic participants who did not undergo diagnostic testing, 1,061 would have tested positive. Consequently, the ability to differentiate between COVID-19-positive and COVID-19-negative cases based on symptoms and sensor data may change over time as testing increases, and as other upper respiratory illnesses such as seasonal influenza increase this fall.

The early identification of symptomatic and pre-symptomatic infected individuals would be especially valuable as transmission is common and people may potentially be even more infectious during this period^{19–21}. Even when individuals have no symptoms, there is evidence that the majority have lung injury (according to computed tomography (CT) scans), and a large number have abnormalities in inflammatory markers, blood cell counts and liver enzymes^{18,22–24}. As the depth and diversity of data types from personal sensors continue to expand—such as heart rate variability (HRV), respiratory rate, temperature, oxygen saturation and even continuous blood pressure, cardiac output and systemic vascular resistance—the ability to detect subtle individual changes in response to early infectious insults will potentially improve and enable the identification of individuals without symptoms.

In the past, the normality of a specific biometric parameter, such as RHR, duration of nightly sleep or daily activity, was based on population norms. For example, a normal RHR is generally considered anything between ~60–100 b.p.m. However, recent work looking at individual daily RHRs over two years found that each person has a

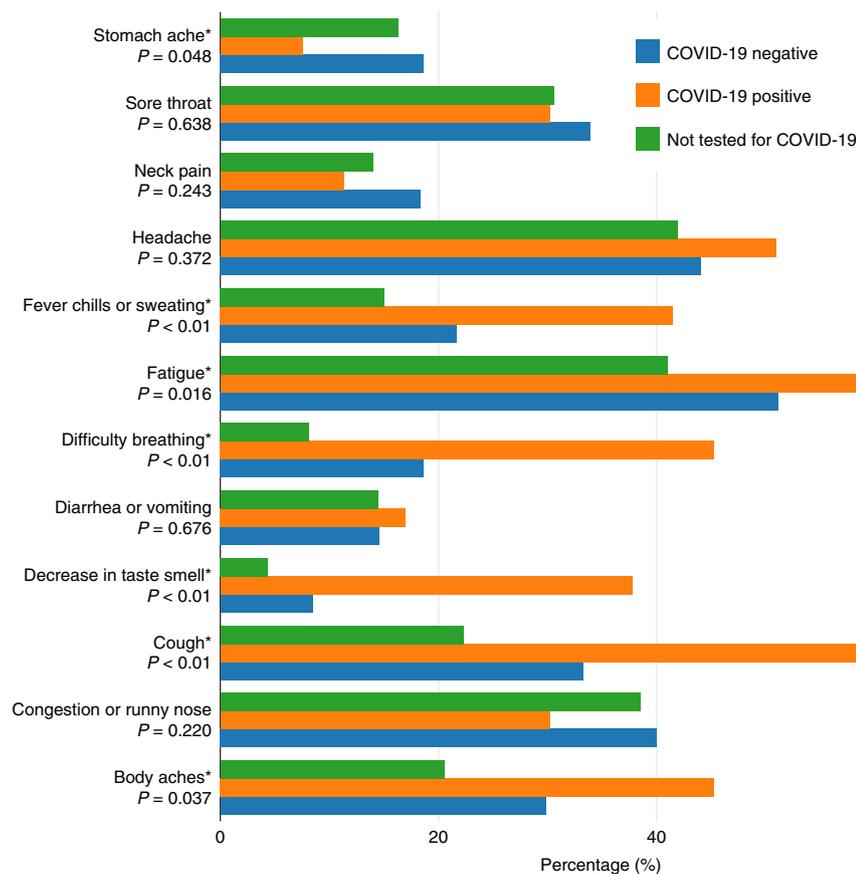


Fig. 1 | Frequency of symptoms among participants. Participants who reported at least one symptom were divided into three groups: participants who tested negative for COVID-19 or positive for COVID-19 and participants who were not tested. The frequencies of the indicated symptoms in each of these three groups are shown. *P* values of a two-sided Fisher's exact test applied to COVID-19-positive (54 individual subjects) and COVID-19-negative (279 individual subjects) participants are reported. Symptoms with a significant difference ($P < 0.05$) are marked with an asterisk.

relatively consistent RHR, for them, that fluctuates by a median of only 3 b.p.m. weekly⁶. On the other hand, what would be considered a normal RHR for an individual can vary by as much as 70 b.p.m. (between 40 and 109 b.p.m.) between individuals. The potential value in identifying important changes in an individual's RHR as an early marker for COVID-19 infection is suggested by the description of 5,700 patients hospitalized with COVID-19⁴: at the time of admission, a greater percentage of individuals had a heart rate of >100 b.p.m. (43.1%) than had a fever (30.7%). Similarly, work in primate models of other viral and bacterial infections found that a significant increase in heart rate can be detected ~2 days before a fever²⁵.

Just as individuals have heart rate patterns that are unique to them, the same is true for sleep patterns. Although population norms for sleep duration have been defined by one-time survey data²⁶, longitudinal analysis of daily sleep over several years supports much greater variation in what is normal for a specific individual⁷. Recognizing what is normal for an individual enables much earlier detection of deviations from that normal.

A strategy of test, trace and isolate has played a central role in helping control the spread of COVID-19. However, testing comes with many challenges, including the enormous logistical and cost hurdles of recurrently testing asymptomatic individuals. In addition, testing in a population with very low prevalence can lead to a high proportion of false positive cases. A refined predictive model, based on personal sensors, could enable an early, individualized testing strategy to improve performance and lower costs. Early testing may make the use of a contact tracing app more effective by identifying positive cases in advance and allowing for early isolation.

DETECT (and similar studies) also represent the transitioning of research from a dependence on brick and mortar research centers to a remote, direct-to-participant approach now possible through a range of digital technologies, including an ever expanding collection of sensors, applications of machine learning to massive datasets, and the ubiquitous connectivity that enables rapid two-way communications 24/7^{27,28}. The promise of digital technologies is that their evolution will continue to bring us closer to identifying the best combination of measures and associated algorithms that identify infection with SARS-CoV-2 or other pathogens. However, it is equally critical to develop and continuously improve on an engaging digital platform that provides value to participants and researchers. This has proven to be extremely challenging, with a recent analysis of eight different digital research programs involving 100,000 participants having a median duration of retention of only 5.5 days²⁹. Digital trials such as DETECT also do come with unique challenges to assure privacy and security, which can only be dealt with by effectively informing participants before consent, storing the data with the appropriate level of security and providing access to the data only for research purposes³⁰. App-based contact tracing, which is not part of DETECT, is an especially sensitive and ethically complicated use of digital technology that can be used to address the pandemic³¹.

Our analyses are dependent entirely on participant-reported symptoms and testing results, as well as the biometric data from their personal devices. Although this is not consistent with the historically more common direct collection of information in a controlled laboratory setting or via electronic health records, previous

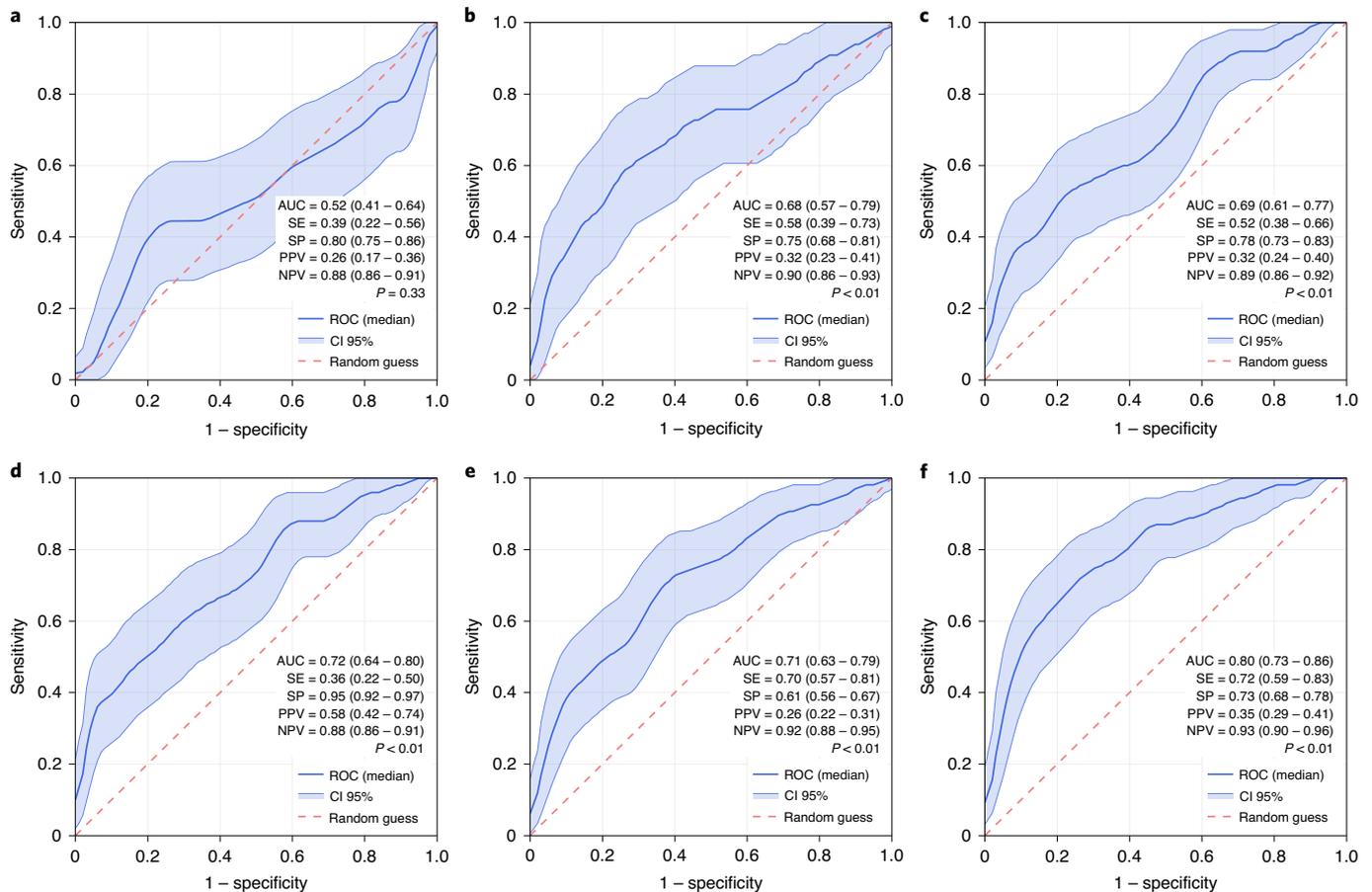


Fig. 2 | Prediction of COVID-19 from self-reported symptoms and sensor data. a–f, Receiver operating characteristic curves (ROCs) for the discrimination between COVID-19-positive (54 individuals) and COVID-19-negative (279 individuals) cases based on the available data: RHR data (**a**); sleep data (**b**); activity data (**c**); all available sensor data (**d**); symptoms only (**e**); symptoms with sensor data (**f**). Models are based on a single decision threshold. Median values and 95% confidence intervals (CIs) for sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV) are reported, considering the point on the ROC with the highest average value of sensitivity and specificity. Error bars represent 95% CIs. P values from a one-sided Mann-Whitney U test are reported.

work has confirmed their value and their accuracy beyond data routinely captured during routine care^{32–34}. Additionally, individuals owning a smartwatch or activity tracker and having access to COVID-19 diagnostic testing are unlikely to be representative of the general population and may exclude those most affected by COVID. Although a recent survey found no racial or ethnic variation in smartwatch or activity tracker usage (23%, 26% and 21% for Black, Hispanic and White individuals, respectively), the lowest percentage of users were identified in those with the lowest annual earnings (12%), the lowest educational attainment (15%) and in those over age 50 (17%)⁵. In the future, if the value of wearable devices to improve individual health is confirmed, this gap in usage will need to be proactively addressed to assure health equity. The decreasing cost of these devices, some now less than US\$35, will help decrease the financial barriers to accomplishing this. Finally, in the early version of the DETECT app we were not able to track the duration or trajectory of individual symptoms, care received and eventual outcomes.

These results suggest that sensor data can incrementally improve symptom-only-based models to differentiate between COVID-19-positive and COVID-19-negative symptomatic individuals, with the potential to enhance our ability to identify a cluster before more spread occurs. Such a passive monitoring strategy may be complementary to virus testing, which is generally a one-off, or infrequent, sampling assay.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-020-1123-x>.

Received: 30 June 2020; Accepted: 7 October 2020;
Published online: 29 October 2020

References

- Menni, C. et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* **26**, 1037–1040 (2020).
- Oran, D. P. & Topol, E. J. Prevalence of asymptomatic SARS-CoV-2 infection. *Ann. Intern. Med.* <https://doi.org/10.7326/M20-3012> (2020).
- New COVID-19 Test Data (Color Genomics, 2020); <https://www.color.com/new-covid-19-test-data-majority-of-people-who-test-positive-for-covid-19-have-mild-symptoms-or-are-asymptomatic>
- Richardson, S. et al. Presenting characteristics, comorbidities and outcomes among 5,700 patients hospitalized with COVID-19 in the New York City area. *JAMA* **323**, 2052–2059 (2020).
- Vogels, E. A. *About One-in-Five Americans Use a Smart Watch or Fitness Tracker* (Pew Research Center, 2020); <https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/>

6. Quer, G., Gouda, P., Galarnyk, M., Topol, E. J. & Steinhubl, S. R. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI and time of year: retrospective, longitudinal cohort study of 92,457 adults. *PLoS ONE* **15**, e0227709 (2020).
7. Jaiswal, S. J. et al. Association of sleep duration and variability with body mass index: Sleep measurements in a large US population of wearable sensor users. *JAMA Intern. Med.* <https://doi.org/10.1001/jamainternmed.2020.2834> (2020).
8. Radin, J. M., Wineinger, N. E., Topol, E. J. & Steinhubl, S. R. Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study. *Lancet Digit. Health* **2**, e85–e93 (2020).
9. Zhu, G. et al. Learning from large-scale wearable device data for predicting epidemic trend of COVID-19. *Discrete Dynamics Nat. Soc.* **2020**, 6152041 (2020).
10. Mishra, T. et al. Early detection of COVID-19 using a smartwatch. Preprint at medRxiv <https://doi.org/10.1101/2020.07.06.20147512> (2020).
11. Natarajan, A., Su, H.-W. & Heneghan, C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. Preprint at <https://doi.org/10.1101/2020.08.14.20175265> (2020).
12. *Evidation Health and BARDA Partner on Early Warning System for COVID-19* (Evidation, 2020); <https://evidation.com/news/evidationhealthandbardapartner/>
13. *Tempredict Study* (Oura Health, 2020); <https://ouraring.com/ucsf-tempredict-study>
14. *Covididentify* (Duke University, 2020); <https://covididentify.covid19.duke.edu/>
15. *Corona Datenspende* (Robert Koch Institut, 2020); <https://corona-datenspende.de/science/en>
16. Shen, B. et al. Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72 (2020).
17. Sharma, R., Agarwal, M., Gupta, M., Somendra, S. & Saxena, S. K. in *Coronavirus Disease 2019 (COVID-19)* (ed. Saxena, S.) 55–70 (Springer, 2020).
18. Tabata, S. et al. Clinical characteristics of COVID-19 in 104 people with SARS-CoV-2 infection on the Diamond Princess cruise ship: a retrospective analysis. *Lancet Infect. Dis.* **20**, 1043–1050 (2020).
19. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
20. Chau, N. V. V. et al. The natural history and transmission potential of asymptomatic SARS-CoV-2 infection. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciaa711> (2020).
21. Jing, Q. L. et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 1141–1150 (2020).
22. Meng, H. et al. CT imaging and clinical course of asymptomatic cases with COVID-19 pneumonia at admission in Wuhan, China. *J. Infect.* **81**, e33–e39 (2020).
23. Inui, S. et al. Chest CT findings in cases from the cruise ship ‘Diamond Princess’ with coronavirus disease 2019 (COVID-19). *Radiol. Cardiothorac. Imaging* **2**, e200110 (2020).
24. Long, Q. X. et al. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nat. Med.* **26**, 1200–1204 (2020).
25. Milechin, L. et al. Detecting pathogen exposure during the non-symptomatic incubation period using physiological data. Preprint at *bioRxiv* <https://doi.org/10.1101/218818> (2017).
26. *Sleep and Sleep Disorders* (Center for Disease Control and Prevention, 2020); https://www.cdc.gov/sleep/data_statistics.html
27. Steinhubl, S. R., Wolff-Hughes, D. L., Nilsen, W., Iturriaga, E. & Califf, R. M. Digital clinical trials: creating a vision for the future. *NPJ Digit. Med.* **2**, 126 (2019).
28. Steinhubl, S. R., McGovern, P., Dylan, J. & Topol, E. J. The digitised clinical trial. *Lancet* **390**, 2135 (2017).
29. Pratap, A. et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *NPJ Digit. Med.* **3**, 21 (2020).
30. Coravos, A. et al. Modernizing and designing evaluation frameworks for connected sensor technologies in medicine. *NPJ Digit. Med.* **3**, 37 (2020).
31. Bradford, L. R., Aboy, M. & Liddell, K. COVID-19 contact tracing Apps: a stress test for privacy, the GDPR and data protection regimes. *J. Law Biosci.* <https://doi.org/10.1093/jlb/ljaa034> (2020).
32. Rivera, S. C. et al. The impact of patient-reported outcome (PRO) data from clinical trials: a systematic review and critical analysis. *Health Qual. Life Outcomes* **17**, 156 (2019).
33. Basch, E. et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. *JAMA* **318**, 197–198 (2017).
34. Bell, S. K. et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. *JAMA Netw. Open* **3**, e205867 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020

Methods

Study population. Any person living in the United States over the age of 18 years old is eligible to participate in the DETECT study by downloading the iOS or Android research app, MyDataHelps. After consenting into the study, participants are asked to share their personal device data (including historical data collected prior to enrollment), report symptoms and diagnostic test results, and connect their electronic health records. Participants can opt to share as much or as little data as they like. Data can be pulled in via direct application programming interface (API) with Fitbit devices, and any device connected through Apple HealthKit or Google Fit data aggregators. Participants were recruited via the study website (www.detectstudy.org), media reports and outreach from our partners at Fitbit, Walgreens, CVS/Aetna and others.

Ethical considerations. The protocol for this study was reviewed and approved by the Scripps Office for the Protection of Research Subjects (IRB 20–7531). All individuals participating in the study provided informed consent electronically.

Statistical analysis. Only participants with self-reported symptoms and COVID-19 test results were considered in this analysis. For each participant, two sets of data were extracted: the baseline data, which included signals spanning from 21 to 7 days before the reported start date of symptoms, and the test data, which included signals beginning at the first date of symptoms to seven days after symptoms. Three types of data were considered from personal sensors: daily resting heart rate (DailyRHR), sleep duration in minutes (DailySleep) and activity based on daily total step count (DailyActivity). The daily resting heart rate is calculated by the specific device³⁵. The total amount of sleep for a given day was based on the total period of sleep between 12 noon of the current day to 12 noon of the next day. When multiple devices from the same individual provided the same information, Fitbit device data were prioritized, for consistency. Overlapping data were combined minute by minute, before aggregating for the whole day.

A single baseline value per individual was extracted for each data type by considering the median value over the individual's baseline data. This value is representative of a participant's 'normal' before the reported symptoms. The baseline value was compared to the test data as follows:

$$\text{RHRMetric} = \frac{\max(\text{DailyRHR}[\text{test data}]) - \text{median}(\text{DailyRHR}[\text{baseline data}])}{4.00}$$

$$\text{SleepMetric} = \frac{\text{mean}(\text{DailySleep}[\text{test data}]) - \text{median}(\text{DailySleep}[\text{baseline data}])}{56.06}$$

$$\text{ActivityMetric} = \frac{\text{mean}(\text{DailyActivity}[\text{test data}]) - \text{median}(\text{DailyActivity}[\text{baseline data}])}{2,489.85}$$

Values were normalized to have a unitary IQR using normalization parameters calculated on all data recorded. For all these metrics, values close to zero indicate small variations from baseline values. This allows us to focus on intra-individual changes, which are minimally affected by the inter-individual variability due to the specific sensor's hardware and estimation algorithms. For the metric based on symptoms only, we adapted the results from the study by Menni et al.¹ to our available data:

$$\text{SymptomMetric} = -1.32 - (0.01 \times \text{age}) + (0.44 \times \text{gender}(\text{male} = 1; \text{female} = 0)) + (1.75 \times \text{DecreaseInTasteSmell}) + (0.31 \times \text{Cough}) + (0.49 \times \text{Fatigue})$$

The multivariate logistic regression model from Menni et al. combined symptoms, age and gender to predict an infection. The parameters were optimized by the authors on a large dataset including over 2 million people, 18,401 of which had undergone a COVID-19 test.

A simple manual metric aggregation strategy without optimization was used to enable a clear understanding of the benefits provided when data from multiple sources were considered together. The aggregated metrics were

$$\text{SensorMetric} = \text{RHRMetric}/10 + \text{SleepMetric} - \text{ActivityMetric}$$

$$\text{OverallMetric} = \text{SensorMetric} + \text{SymptomMetric}$$

The main outcomes are ROC curves for each of the proposed metrics. The curves are obtained by considering a binary classification task between participants self-reported as COVID-19-positive and COVID-19-negative. The models are based on a single decision threshold, which is directly compared to the metric values, with the aim of minimizing overfitting issues while providing a fair comparison. Confidence intervals, reported with a confidence level of 95%, are estimated using a bootstrap method by repeatedly sampling the dataset with replacement. The sampling is performed in a stratified manner; that is, the balance of the classes is maintained over all experiments. Values for sensitivity (SE), specificity (SP), positive predictive value (PPV) and negative predictive value (NPV) were also calculated (Fig. 2). SE and SP are defined as the fraction of positive and negative individuals correctly classified, respectively, while PPV and NPV are the fraction of individuals predicted as positive and negative that are correctly classified, respectively. These values are based on the point in the ROC with the optimal tradeoff between sensitivity and specificity, which may vary depending on the shape of the curve. For each metric analyzed, we applied the one-sided Mann–Whitney *U* test with the alternate hypothesis that the underlying model of the positive class is stochastically greater than the negative class. All statistical tests were evaluated using the Python package *scipy* version 1.5.2. The comparison metric to assess the overall performance was the AUC of the ROC.

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

Data availability

All interested investigators will be allowed access to the analysis dataset following registration and pledging to not re-identify individuals or share the data with a third party. All data inquiries should be addressed to the corresponding author.

References

35. Heneghan, C., Venkatraman, S. & Russell, A. Investigation of an estimate of daily resting heart rate using a consumer wearable device. Preprint at *medRxiv* <https://doi.org/10.1101/19008771> (2019).

Acknowledgements

This work was funded by grant no. UL1TR002550 from the National Center for Advancing Translational Sciences (NCATS) at the National Institutes of Health (NIH; E.J.T. and S.R.S.). We thank N. Dalton for his support of DETECT. We also thank D. Oran, T. Peters, R. Kamyar and C. Nowak for their contributions to DETECT.

Author contributions

G.Q., J.M.R., K.B.-M., V.K., E.J.T. and S.R.S. made substantial contributions to the study conception and design. K.B.-M., L.A., E.R., V.K. and S.R.S. made substantial contributions to the acquisition of data. G.Q. and M.G. conducted statistical analysis. G.Q., J.M.R., M.G. and S.R.S. made substantial contributions to the interpretation of data. G.Q., J.M.R. and S.R.S. drafted the first version of the manuscript. G.Q., J.M.R., M.G., K.B.-M., L.A., E.R., V.K., E.J.T. and S.R.S. contributed to critical revisions and approved the final version of the manuscript. G.Q., J.M.R. and S.R.S. take responsibility for the integrity of the work.

Competing interests

S.R.S. reports grants from Janssen and personal fees from Otsuka and Livongo, outside the submitted work. The other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41591-020-1123-x>.

Correspondence and requests for materials should be addressed to G.Q.

Reprints and permissions information is available at www.nature.com/reprints.

Editor recognition statement Michael Basson was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

The MyDataHelps smartphone based app was developed by CareEvolution, and it includes the DETECT study since March 25, 2020. The participants have agreed to share the historical data collected by their eligible devices, including heartrate, sleep and activity measures. On first use, the app records self-reported age, gender and location. With continued use the users are encouraged to share additional information through the completion of targeted surveys, including manifestation of symptoms and COVID-19 test outcomes. Data from the participants is automatically uploaded to a protected server.

Data analysis

Analyses were carried out using Python version 3.8.5. The Python packages pandas version 1.1.2 and numpy version 1.19.1 have been used for data processing. Receiver operating characteristic (ROC) curves were computed using the Python package scikit-learn version 0.23.2. Statistical tests and p-values have been evaluated using the Python package scipy version 1.5.2.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All interested investigators will be allowed access to the analysis data set following registration and pledging to not re-identify individuals or share the data with a third party. All data inquiries should be addressed to the corresponding author.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Between March 25, 2020 and June 7, 2020, 30,529 US individuals were enrolled in the DETECT study. During the study, 3,811 individuals reported symptoms through the DETECT smartphone app. 480 individuals reported having had a SARS-CoV-2 test, 333 of which having received the outcome (54 individuals tested positive; 279 tested negative). We applied bootstrap resampling method with 10,000 independent iterations to estimate the uncertainty of the outcomes due to the sample size. Due to the observational nature of the study, we did not perform any statistical analysis to predetermine the sample size, which was enforced by the number of active participants enrolled.
Data exclusions	The 3,811 individuals who completed the report for symptoms in the DETECT app were included in the analysis of frequency of symptoms. All the 333 individuals who had received the outcome of the SARS-CoV-2 test were included in the analysis of the proposed models. The data exclusion criteria for the analysis were pre-established. Due to the nature of the study, only individuals owning a smartwatch or activity tracker device were able to be enrolled. No data samples that were eligible for the analysis has been excluded.
Replication	The study has not been replicated, due to the nature of the data collection process, requiring the enrollment of thousands of individuals. The methods have been described in detail in the paper with all the software packages adopted, so it is possible to replicate the study by accessing a similar but independent dataset.
Randomization	Randomization was not relevant for this study, as participants have not been divided in two or more cohorts at the beginning of the study. The study focuses on the discrimination between individuals who tested positive or negative to COVID-19, so no randomization is needed.
Blinding	The group allocation was solely related to the data shared by the participants. Due to the observational nature of the study, the investigators were not involved in any arbitrary group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Between March 25, 2020 and June 7, 2020, our research study enrolled 30,529 individuals with representation from every state in the United States. Among the consented individuals, 62.0% are female and 12.8% are 65 or more years old. 78.4% of participants connected their Fitbit device to the study-app, 31.2% connected the data from Apple Health Kit while 8.1% connected data from Google Fit (note that one individual can connect to multiple platforms). The most important covariates available at the time of analysis are sex, age and device used by each participants. We did not observe any statistically significant difference among subjects with reported positive and negative COVID-19 test.
Recruitment	Any person living in the United States over the age of 18 years old is eligible to participate in the DETECT study by downloading the iOS or Android research app, MyDataHelps. Scripps Research, along with outreach partners, conducted a multi-faceted outreach campaign including social media, educational web-based content, email outreach and in-app notifications to recruit participants for the DETECT Study. The study was also featured in several national media outlets, which increased visibility to larger section of the U.S. population.

Individuals owning a smartwatch or activity tracker and having access to COVID-19 diagnostic testing may not be fully representative of the general population.

Ethics oversight

The protocol for this study was reviewed and approved by the Scripps Office for the Protection of Research Subjects (IRB 20-7531). All individuals participating in the study provided informed consent electronically.

Note that full information on the approval of the study protocol must also be provided in the manuscript.