

## COMPUTATIONAL MEDICINE

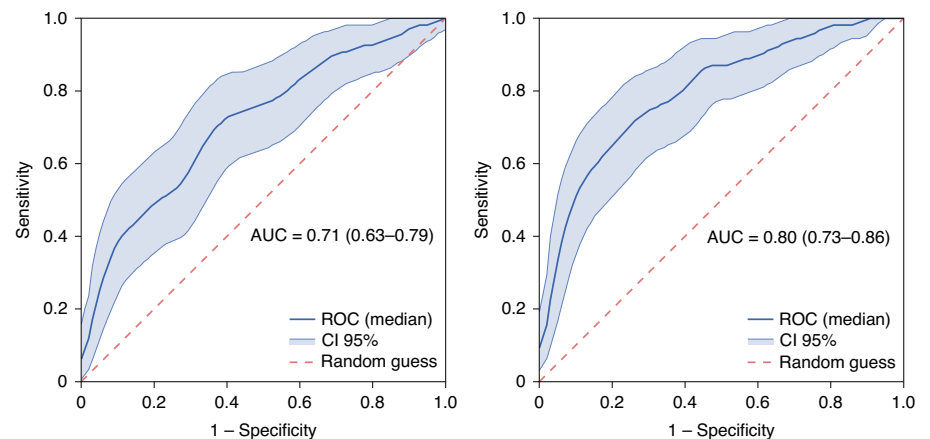
## Smartwatch data help detect COVID-19

Data from consumer smartwatches can improve the detection of COVID-19 when combined with symptom self-reporting, and can also detect the disease in pre-symptomatic individuals.

Tingting Zhu, Peter Watkinson and David A. Clifton

Tests for the detection of COVID-19 are typically time consuming, costly and require professional expertise. Improving the frequency, ease and ubiquity of testing for COVID-19 is urgent, particularly when a substantial proportion of patients (40–45%; ref. <sup>1</sup>) may be pre-symptomatic or asymptomatic. Obtaining longitudinal physiological data via commonplace wearable devices<sup>2</sup>, typically worn on the wrist, may offer a convenient means of detection. Self-reported symptoms can be used to construct relatively simple models for the identification of COVID-19 (ref. <sup>3</sup>), and data from wearables may similarly be used to identify viral respiratory illnesses<sup>4,5</sup>. Reporting in *Nature Medicine*, Giorgio Quer and colleagues now show how smartwatch data can be used in conjunction with self-reported symptoms to determine whether an individual has COVID-19 after the onset of symptoms<sup>6</sup>. And in *Nature Biomedical Engineering*, Michael Snyder, Xiao Li and colleagues report how similar data, also from consumer smartwatches, can be used in advance of symptom onset to identify, and potentially predict, COVID-19 infection<sup>7</sup>.

Between March–June 2020, Quer and co-authors conducted the DETECT study, in which 30,529 participants in the United States provided data from their smartwatches and activity trackers (78.4% of the participants used Fitbit devices, 31.2% used the Apple Watch, and 8.1% used devices compatible with Google Fit; some participants used more than one platform). Some of the participants also self-reported symptoms (3,811 participants, or 12.4% of the total) and the results of diagnostic tests (333 participants, or 8.7%; 54 participants, or 16.2%, reported a positive test result). For those with test results, the authors analysed their daily average resting heart rate (RHR; in beats per minute), daily sleep duration (in minutes) and daily activity (step count) according to two intervals: a baseline window of 7–21 days before the onset of symptoms, and a ‘test interval’ spanning 7 days from symptom onset. For each participant and data type, the

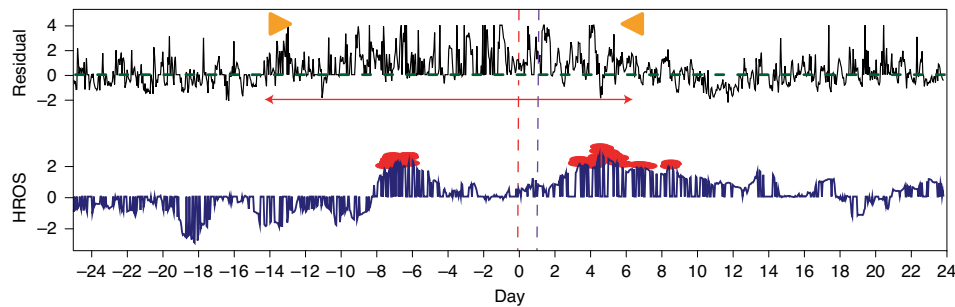


**Fig. 1 | Prediction of COVID-19 from self-reported symptoms, and from self-reported symptoms combined with RHR, sleep and activity data from smartwatches.** The receiver operating characteristic (ROC) curves for the discrimination of 54 individuals who tested positive for COVID-19 and 279 individuals who tested negative for the disease show an AUC of 0.71 for the symptom-based model (left) and of 0.80 for the model using symptoms and smartwatch data (right). CI, 95% confidence interval. Figure reproduced with permission from ref. <sup>6</sup>, Springer Nature Ltd.

authors calculated the differences between the maximum values (for RHR) or mean values (for sleep and activity data) in the test interval and the median of the values in the baseline window. The differences were then combined in a number of heuristic metrics, which were used to classify each participant as COVID-19-positive or COVID-19-negative. When compared to the self-reported results from the diagnostic tests (considered as ground truth), a metric aggregating the smartwatch data and the self-reported symptoms led to an area under the receiver operating characteristic curve (AUC) of 0.80; an existing heuristic model<sup>3</sup> that uses symptoms alone led to an AUC of 0.71 (Fig. 1).

Snyder and co-authors used a dataset collected in February–June 2020 from 5,262 participants who completed surveys related to respiratory illness, symptoms and diagnosis, and who wore Fitbit devices (63.2%), the Apple Watch (18.7%) or Garmin smartwatches (8.1%). One hundred and fourteen participants (2.2%) had COVID-19 and provided symptoms

and diagnosis dates; 47 participants (0.9%) provided symptoms and diagnosis dates for different respiratory infections. The authors analysed a subset of 32 participants with a positive COVID-19 test (0.6%), for whom sufficient smartwatch data (RHR, sleep duration and step counts) spanning the interval of infection to disease (including symptom dates and diagnosis dates) were available, as well as 73 healthy participants and 15 participants who had other respiratory illnesses. Variations in sleep duration before and after symptom onset were examined, but sleep durations were not used as input. The authors constructed and assessed three models, involving differences between the observed RHR and the average RHR over a sliding window of 28 days; a statistic on cumulative sums of RHR, which the authors also suggest can be used to alert the user of potential infection in real time; and an anomaly-detection method involving the ratio of heart rate and daily steps. For the first two models, alerts were defined via signal thresholding; for the latter model, alerts were determined via binary



**Fig. 2 | Heart-rate metrics for an individual before COVID-19 infection and during illness.** The red dashed line indicates the day of symptom onset and the purple dashed line the date of diagnosis. Top: RHR residuals (with respect to the average RHR within a 28-day window. The green dashed line lies at zero (null residuals). The gold triangles denote the window of infection detection, according to a parametrized model. The red double-headed arrow indicates significantly elevated residuals. Bottom: smoothed heart rate over steps (HROS), normalized according to a Gaussian distribution of data from 1-h intervals. The red dots indicate anomalies in the normalized HROS. Figure reproduced with permission from ref. <sup>7</sup>, Springer Nature Ltd.

classification (normal versus anomalous). The authors defined a detection window starting 14 days before symptom onset and ending 7 days after, and compared the alerts raised by the three methods with the reported symptom dates and diagnosis dates. They found that 22 of the 32 participants with COVID-19 would have received alerts ahead of symptom onset or in the same day, regardless of the method, and that 15 out of the 24 participants for whom more than 28 days of smartwatch data were available would have received an alert from the method using cumulative sums of RHR on or before symptom onset (Fig. 2).

Obtaining sufficient and useful data is difficult. In both studies, the number of participants who downloaded a smartphone app, successfully linked it to their smartwatch, were tested for COVID-19 and reported the result (positive or negative), reported any symptoms, and used their smartwatch for sufficiently long periods, is small. In the study by Quer and colleagues, 1.1% of all participants had sufficient data for analysis, and 0.2% of all participants reported a positive COVID-19 test result; in Snyder and colleagues' study, 2.2% of all participants were analysed, and 0.6% of all participants had a positive COVID-19 test and had their data included in the analysis. Hence, population-scale analyses of wearable data for disease detection will need to be designed with sufficient robustness for practical applications. The use of reporting guidelines for prediction models (such as transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)) would facilitate comparisons among studies. Importantly, as noted by Quer and co-authors, selection bias inherent to the use of smartwatches should be considered; in fact, 87.2% of the participants in the authors' study were younger than 65, and smartwatch ownership

is low in populations that are most at risk for COVID-19, such as low-income groups<sup>2</sup>. Other inclusion biases relate to the reporting of the results of COVID-19 tests, which in most countries are preferentially given to individuals with serious symptoms. Similarly, any direct extrapolation of true-positive rates from these types of studies to wider populations would involve the unlikely assumption of equivalent distributions. Some of these potential biases could be overcome by systems that collect data across a wide range of diverse users. At the expense of accuracy, sleep duration and activity levels that are self-reported (rather than obtained from wearables) may widen study participation and reduce inclusion bias.

Performing online repeated analysis of physiological time series to predict rare events involves further complexities, such as how to evaluate multiple-testing results (that is, day-by-day predictions) for each participant. Quer and colleagues sidestep this issue by making one binary classification per patient (all classifications are then independent because each is associated with an independent participant, and each participant has either a positive or negative COVID-19 test). Yet evaluating time series models, as in Snyder and colleagues' study, is trickier: the models output a series of alerts, which may or may not align with particular events (such as symptom onset) and intervals (such as pre-symptomatic periods or prodromal periods). Such 'repeat' classifications are arguably neither binary nor independent, as recordings from an individual may include data from healthy periods and disease periods.

Model complexity is typically a concern in the analysis of physiological time series. Quer and colleagues defined straightforward heuristics to avoid overfitting the model

to the data, yet it is unclear whether the parameters used within the heuristics were selected so as to maximize performance in the same dataset (which could be considered 'in-sample testing'). With Snyder and colleagues' models, cross-validation may minimize the possibility of overfitting the models to the available data. When working with small datasets, poor generalization to unseen datasets is a possibility.

Data from wearables are particularly prone to noise and artefacts. Snyder and co-authors show that both healthy participants and participants who had COVID-19 could have received 'false positive' alerts from the models. The authors hypothesize that some of these alerts might be caused by end-of-year holidays and other similar events. Sufficiently large studies should offer the possibility of obtaining data from non-disease events such that false positives can be reduced. However, some of these alerts could actually be true positives, as even gold-standard diagnostic tests can fail to detect COVID-19.

When data availability is not a constraint, complexity may be added to the models by increasing the amount of input data beyond the daily metrics produced by wearable devices, for instance by incorporating features derived from the underlying accelerometry and photoplethysmography waveforms that many of these devices acquire. In fact, respiratory rate and cardiovascular parameters can be estimated from such signals<sup>8,9</sup>. Additionally, machine-learning techniques could be used for the automated discovery of features that could be incorporated in a predictive model based on multiple time-varying physiological variables<sup>10</sup>. The studies of Quer, Snyder and their respective colleagues suggest that one day devices on our wrists could accurately alert us to a potential infection before we get sick. □

Tingting Zhu<sup>1</sup>, Peter Watkinson<sup>2</sup> and David A. Clifton<sup>1</sup>✉

<sup>1</sup>Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK. <sup>2</sup>Kadoorie Centre for Critical Care Research, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK.

✉e-mail: [david.clifton@eng.ox.ac.uk](mailto:david.clifton@eng.ox.ac.uk)

Published online: 8 December 2020

<https://doi.org/10.1038/s41551-020-00659-9>

#### References

1. Oran, D. P. & Topol, E. J. *Ann. Intern. Med.* **173**, 362–367 (2020).
2. Vogels, E. A. About one-in-five Americans use a smart watch or fitness tracker. *Pew Research Center* <https://pewrsr.ch/37TaaN4> (2020).
3. Menni, C. et al. *Nat. Med.* **26**, 1037–1040 (2020).
4. Zhu, G. et al. *Discrete Dyn. Nat. Soc.* **2020**, 6152041 (2020).
5. Li, X. et al. *PLoS Biol.* **15**, e2001402 (2017).
6. Quer, G. et al. *Nat. Med.* <https://doi.org/10.1038/s41591-020-1123-x> (2020).
7. Mishra, T. et al. *Nat. Biomed. Eng.* <https://doi.org/10.1038/s41551-020-00640-6> (2020).
8. Jarchi, D. et al. *Healthc. Technol. Lett.* **6**, 19–26 (2019).
9. Charlton, P. H. et al. *IEEE Rev. Biomed. Eng.* **11**, 2–20 (2018).
10. Shamout, F. E. et al. *IEEE J. Biomed. Health Inform.* **24**, 437–446 (2020).

#### Competing interests

D.A.C. has received grants from GlaxoSmithKline and personal fees from Oxford University Innovation, Biobeats and Sensyne Health, for work unrelated to this discussion. P.W. holds grants from the NIHR, Wellcome, and Sensyne Health, for work unrelated to this discussion. He was Chief Medical Officer for Sensyne Health, and holds shares in the company. T.Z. declares no competing interests.