# The performance of wearable sensors in the detection of SARS-CoV-2 infection: a systematic review

Marianna Mitratza*, Brianna Mae Goodale*, Aizhan Shagadatova, Vladimir Kovacevic, Janneke van de Wijgert, Timo B Brakenhoff, Richard Dobson, Billy Franks, Duco Veen, Amos A Folarin, Pieter Stolk, Diederick E Grobbee, Maureen Cronin, George S Downward

**Containing the COVID-19 pandemic requires rapidly identifying infected individuals. Subtle changes in physiological parameters (such as heart rate, respiratory rate, and skin temperature), discernible by wearable devices, could act as early digital biomarkers of infections. Our primary objective was to assess the performance of statistical and algorithmic models using data from wearable devices to detect deviations compatible with a SARS-CoV-2 infection. We searched MEDLINE, Embase, Web of Science, the Cochrane Central Register of Controlled Trials (known as CENTRAL), International Clinical Trials Registry Platform, and ClinicalTrials.gov on July 27, 2021 for publications, preprints, and study protocols describing the use of wearable devices to identify a SARS-CoV-2 infection. Of 3196 records identified and screened, 12 articles and 12 study protocols were analysed. Most included articles had a moderate risk of bias, as per the National Institute of Health Quality Assessment Tool for Observational and Cross-Sectional Studies. The accuracy of algorithmic models to detect SARS-CoV-2 infection varied greatly (area under the curve 0·52–0·92). An algorithm's ability to detect presymptomatic infection varied greatly (from 20% to 88% of cases), from 14 days to 1 day before symptom onset. Increased heart rate was most frequently associated with SARS-CoV-2 infection, along with increased skin temperature and respiratory rate. All 12 protocols described prospective studies that had yet to be completed or to publish their results, including two randomised controlled trials. The evidence surrounding wearable devices in the early detection of SARS-CoV-2 infection is still in an early stage, with a limited overall number of studies identified. However, these studies show promise for the early detection of SARS-CoV-2 infection. Large prospective, and preferably controlled, studies recruiting and retaining larger and more diverse populations are needed to provide further evidence.**

*Lancet Digit Health* 2022; 4: e370–83

*Joint first authors

**Julius Global Health** (M Mitratza PhD, A Shagadatova MSc, D Veen PhD, Prof D E Grobbee PhD, G S Downward PhD), **Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands** (Prof J van de Wijgert PhD, P Stolk PhD); **Ava AG, Zurich, Switzerland** (B M Goodale PhD, V Kovacevic PhD, M Cronin PhD); **Julius Clinical Research BV, Zeist, Netherlands** (T B Brakenhoff PhD, B Franks PhD, D Veen, Prof D E Grobbee); **Institute of Health Informatics, University College London, London, UK** (Prof R Dobson PhD, A A Folarin PhD); **Optentia Research Program, North-West University, Potchefstroom, South Africa** (D Veen); **National Institute for Health Research Maudsley Biomedical Research Centre, King's College London, London, UK** (A A Folarin); **Department of Biostatistics and Health Informatics, South London and Maudsley NHS Foundation Trust, London, UK** (A A Folarin)

Correspondence to: Dr Marianna Mitratza, Julius Global Health, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht 3508GA, Netherlands **m.mitratza@umcutrecht.nl**

## Introduction

On Dec 31, 2019, WHO recognised the emergence of SARS-CoV-2, a novel virus in the coronavirus family.[1] Since then, the outbreak of illness caused by the SARS-CoV-2 virus (COVID-19) has become a global pandemic, causing more than 458 million cases and 6 million deaths, until March, 2022.[2]

A key strategy for containing the COVID-19 pandemic has been the rapid identification and contact tracing of infected individuals.[3,4] RT-PCR constitutes the gold standard for diagnostic testing of COVID-19.[5–7] Despite developments in rapid testing, the timing of testing in relation to stage of infection hinders public health efforts to control the virus.[8] On average, from SARS-CoV-2 infection to symptom onset takes 6 days, although the incubation period can be as long as 18 days.[9] The viral load from the upper respiratory tract increases during the incubation period, reaches a peak around symptom onset, and then gradually declines.[10] Many national health guidelines recommend testing for the general population after symptom onset, or a few days after suspected exposure to the virus, regardless of symptoms, to limit false-negative test results.[11–14] However, viral load could be sufficiently high enough for transmission before people have symptoms or qualify for testing.[15,16]

COVID-19 remains difficult to distinguish from other respiratory illnesses on the basis of reported symptoms alone. Many common COVID-19 symptoms (eg, fever and cough) overlap with other influenza-like illnesses.[17,18] Some patients with confirmed COVID-19 report symptoms uniquely associated with the virus (eg, anosmia), but such symptoms rarely appear early in the disease.[19] Furthermore, 20–30% of individuals infected with SARS-CoV-2 never develop symptoms.[20–22] The US Centers for Disease Control and Prevention report that presymptomatic or asymptomatic people account for half of SARS-CoV-2 virus transmissions.[23]

To reduce transmission rates in the general population, identifying SARS-CoV-2 infections before or in the absence of symptom onset is crucial. A range of non-invasive, commercially available physiological monitors (ie, wearable devices) could help in detecting presymptomatic and asymptomatic infections and controlling the pandemic. Because of rapid technological advancements, relatively subtle fluctuations in physiological parameters such as body temperature, respiratory rate, heart rate, heart rate variability, skin perfusion, and oxygen saturation ($SpO_2$) can be measured by sensors commonly found in smartwatches, smart rings, and fitness trackers. Fever remains one of the most commonly reported COVID-19 infection symptoms;[24] thus, the inclusion of thermometer sensors on an increasing number of wearable devices, despite their reliance on sensors worn on distal body parts, might render them suitable to detecting SARS-CoV-2 infection. Of note, peripheral temperatures measured by wearable devices have shown greater sensitivity than oral measurements in detecting subtle temperature shifts (eg, ≥0·2°C).[25] With regard to the COVID-19 pandemic, wrist temperatures have been found to be equally stable and less susceptible to environmental influences than forehead temperatures.[26] Calls for additional research on the role wearable devices could serve in the early and comprehensive detection of SARS-CoV-2 infections have emphasised their potential ability to inform population and individual health responses to the pandemic.[27] Several studies, mostly of retrospective design, have shown the feasibility of wearable devices in indicating the presence of SARS-CoV-2 infection by monitoring one or more physiological parameters, but an overview of the evidence is not yet available.

In this systematic review, we aimed to summarise and assess the added value of wearable devices in the detection of SARS-CoV-2 infection within the adult population (ie, those 18 years and older). Our primary question regards the current state of evidence on the diagnostic accuracy of statistical and algorithmic models using wearable sensor data. We also consider the time from detection to symptom onset and which physiological parameters provide the best indication of a subclinical or symptomatic SARS-CoV-2 infection.

## Methods

### Search strategy and selection criteria

We conducted our systematic review in line with our protocol[28] and report our findings according to PRISMA recommendations. We initially searched the literature between Dec 17 and Dec 21, 2020, on the electronic databases PubMed (MEDLINE), Embase, Web of Science, Cochrane Central Register of Controlled Trials (known as CENTRAL), International Clinical Trials Registry Platform, and ClinicalTrials.gov. As the use of wearables to identify SARS-CoV-2 infections remains an ongoing area of research, we also searched preprint repositories (medRxiv and bioRxiv) for non-peer-reviewed studies between Dec 17 and Dec 21, 2020. We manually searched the reference lists of articles and reviews included for full-text screening to identify additional relevant studies. To ensure as current a review as possible, we repeated the above searches on March 8, 2021, and March 9, 2021, and again on July 27, 2021, before final analysis.

See Online for appendix

The search terms for each database (appendix pp 3–5) were selected on the basis of the authors' knowledge regarding wearable devices and SARS-CoV-2 infection. All databases were searched for the years 2020 and 2021, aligning with WHO's timeline of SARS-CoV-2 discovery.[1] We did not restrict our search by setting or language.

Articles and protocols showing randomised controlled trials (RCTs), non-RCTs, and observational studies (prospective and retrospective) were eligible for inclusion, provided they examined wearable devices' detection of SARS-CoV-2 infection in a non-hospitalised population. We defined wearable devices as non-invasive body-worn sensors automatically monitoring one or more physiological parameters in real-time, including—but not limited to—skin temperature, respiratory rate, heart rate, heart rate variability, or skin perfusion or a combination of these parameters. Additional criteria for study selection included reporting on how SARS-CoV-2 was diagnosed (ie, a reference test). Studies reporting on exclusively inpatient or paediatric and adolescent populations (ie, those 17 years and younger), internal wearable devices, wearables requiring manual data collection, or wearables designed for hospital settings were excluded. Case reports, editorials, commentaries, personal opinions, and animal studies were also not eligible for inclusion.[28]

### Data analysis

We provide detailed descriptions of data extraction and analysis in the appendix (p 6). Briefly, all articles found via our search underwent deduplication and title and abstract screening. Two authors (MM and AS for the initial search) then reviewed the full text of all papers identified and included during the initial screening. Any discrepancies were resolved through discussions with a third reviewer (GSD). Papers meeting our inclusion criteria underwent data extraction to obtain study-level information on participant demographics, study design and setting, sample size, the type of wearable device and it's sensors, reference test, definition of key model parameters and features, and performance metrics (eg, area under the curve [AUC] and other test statistics). We contacted all corresponding authors to discuss missing data and areas of uncertainty. Finally, we assessed the risk of bias for each study's primary outcomes using an adapted version of the National Institutes of Health Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies.[29] Per our protocol, a meta-analysis of the results could not be done, given the heterogeneity in approaches and outcomes.

## Results

The first database search, done on Dec 17–21, 2020, identified 1601 records with an additional four articles retrieved from manually screening review reference lists. The second search, conducted on March 8–9, 2021, found an additional 574 records, and the third search, done on July 27, 2021, found an additional 1691 records, resulting in 3196 unique records overall, after deduplication. After title and abstract screening, 173 articles were retained for full-text review, of which 12[19,30–40] fulfilled our inclusion and exclusion criteria (appendix p 7). All studies were observational, and seven were strictly retrospective;[19,30–32,35,38,40] although some researchers implemented control procedures, no RCTs were reported. Our searches also identified 12 study protocols,[41–52] including two RCTs.[43,50] Eight protocols were recorded in online registries; one was a preprint, and three were published (appendix pp 8–9). During extraction, we contacted the corresponding authors for studies with missing data and received replies from six of the 12 research teams.

We compiled the key characteristics for the 12 studies included in this systematic review (table 1; see appendix pp 11–14 for a detailed description). Most studies recruited active users of wearable devices with a

| | Study design/ Population and study setting | Wearable device | COVID-19-positive sample size (n)/ Total analysed sample size (n) | Race or ethnicity n (%) | Reference standard | Timing of the detected deviation compared to SO (days)* | Algorithm or statistical model | Key features in the best performing model | Training set, n (%) | Test, validation (internal or external), or comparison set, n (%) | Key findings for the best performing model |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Bogu and Snyder (2021)[40]†‡ | Observational and retrospective/ Subset of Mishra et al's (2020)[33] data | Fitbit | 25 /106 | NR | Self-reported COVID-19 diagnosis confirmed with a physician note | SO –6·94 to +5·12 | Long-term short-term memory networks-based autoencoder for anomaly detection (known as LAAD) | Changes in resting HR from baseline | NR | NR | Positive predictive value 0·91 (95% CI 0·854–0·967); sensitivity 0·36 (95% CI 0·232–0·487); F-beta (0·1) 0·79 (95% CI 0·693–0·888); abnormal HR lasted longer in the COVID-19-positive cohort; more COVID-19-positive cases had >1 day of abnormal HR |
| Cleary et al (2021)[31]† | First year medical interns, USA/ Observational and retrospective | FitBit Inspire HR and the Apple Charge 3 Watch | 22 /105 | NR | Self-reported SARS-CoV-2 test | NA | Binary classifier | Resting HR, sleep duration (min), and total step count | 105 (100%) | NA | Activity data: AUC 0·75 (95% CI 0·63–0·87); all sensor data: AUC 0·75 (0·62–0·89) |
| Hassantabar et al (2020)[36]† | Observational and cross-sectional/ Health-care workers and patients of San Matteo hospital, Pavia, Italy | Empatica E4, pulse oximeter, and blood pressure monitor | 57/87 | NR | PCR test upon hospital arrival | ND | Deep neural networks | Galvanic skin response, $SpO_2$, blood pressure, and questionnaire data (eg, on symptoms, presence of chronic lung diseases, and whether participants are immuno-compromised) | 52 (60%; 18 healthy, 16 asympto-matic, and 18 sympto-matic) | 17 (20%) in test set (6 healthy, 5 asymptomatic, and 6 symptomatic); 18 (20%) in the validation set (6 healthy, 6 asymptomatic, and 6 symptomatic) | F1 (ie, the harmonic mean of precision and sensitivity) 98·2%; true positive 98·1%; false positive 0·8%; false negative 0% for symptomatic individuals; sensitivity 97·52%; specificity 99·16% |
| Hirten et al (2021)[37] | Observational and prospective/ American health-care workers at Mount Sinai Health System, New York, NY, USA | Apple Watch Series 4 or 5 | 13/297 | 73 Asian (24·6%), 29 Black (9·8%), 43 other (14·5%), 108 White (36·4%), 44 Hispanic ethnicity (14·8%) | Self-reported nasal PCR test | ND | Mixed-effect cosinor model | HRV (SD of normal to normal R–R intervals) including mean MESOR, acrophase, and amplitude | NA | NA | Shorter mean SD of normal to normal R–R intervals amplitude in participants positive for COVID-19 |
| Lonini et al (2021)[39]‡ | Observational, cross-sectional/ American healthy controls or COVID-19-positive patients recovering at home or in a hospital physical rehabilitation centre | Unnamed throat-worn patch | 15/29 | NR | Tested positive for COVID-19 | NR | Logistic regression with elastic net regularisation | HR, HRV (SD of R–R intervals), respiratory rate, cough frequency, and walk cadence | 29 (100%); randomly sampled one walk sequence and one cough sequence with replacement five times per individual, repeated 100 times to estimate CI | Validation set was leave-one-subject-out cross validation | AUC ≥0·92 (95% CI 0·92–0·96); pre-walk HR higher in COVID-19-positive cohort; pre-walk respiratory rate higher in COVID-19-positive cohort; pre-walk HRV lower for COVID-19-positive cohort; COVID-19-positive cohort walk slower |

(Table 1 continues on next page)

(Continued from previous page)

| | Study design / Population and study setting | Wearable device | COVID-19-positive sample/ Total analysed sample size (n/N) | Race or ethnicity n (%) | Reference standard | Timing compared to SO (days)* | Algorithm or statistical model | Key features in the best performing model | Training set, n (%) | Test, validation (internal or external), or comparison set, n (%) | Key findings for the best performing model |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Miller et al (2020)[30] | Observational and retrospective/ Ambulatory, opt-in study of device users | WHOOP | 81/271 | NR | Self-reported SARS-CoV-2 test | ND | Gradient boosted classifier | 5 respiratory rate-derived features | 57 (70%); COVID-19-positive with symptoms between March 14 and April 14, 2020 | 24 in validation set one (30%); COVID-19-positive with symptoms between April 14–June 6, 2020; 190 participants negative for COVID-19 in validation set two | Sensitivity 36·5%, specificity 95·3%, positive predictive value 73·8%, negative predictive value 80·6% for the test set; identified 20% of individuals positive for COVID-19 before SO; identified 80% of individuals positive for COVID-19 by SO +3 days |
| Mishra et al (2020)[33] | Observational, prospective, and retrospective/ Ambulatory, opt-in study of device users | Fitbit (Ionic, Charge 4, and Charge 3) | 32/120 | 27 European (84·4%), 5 mixed or other (15·6%)§ | Self-reported COVID-19 (diagnosis confirmed with physician note) | SO −28 to SO +7¶ | Offline (HROS-AD, RHR-Diff) and online (CuSum) anomaly detection algorithms | The HROS-AD model included HR and step count as features; the RHR-Diff model included HR; the CuSum model included deviations in elevated residual resting HR | 32 (100% of participants positive for COVID-19) | 73 self-reported healthy participants in comparison set one; 15 participants not positive for COVID-19 in comparison set two | Median time to SO from elevated HR was 4 days, median HR increased by 7 beats per min following SO, step count decreased at onset of HR changes associated with COVID-19, sleep duration increased at onset of HR changes associated with COVID-19 when missing data was imputed, CuSum detected 63% of SARS-CoV-2 infections before SO in real-time |
| Natarajan et al (2020)[35] | Observational and retrospective/ Ambulatory, opt-in study of American and Canadian device users | Fitbit | 1257 | NR | Self-reported PCR test | SO −1 to SO +4¶ | Convolutional neural network | Body-mass index, age, sex, mean nocturnal respiratory rate, mean nocturnal HR during non-rapid eye movement sleep, HRV (RMSSD of nocturnal respiratory rate series), Shannon entropy of nocturnal respiratory rate series, and data from the day of examination and the 4 preceding days | 879 (random 70% split); 70:15:15 split performed five times, but cross-validation performed only once | 189 in test set (15%); 189 in cross- validation set (15%) | Sensitivity 25·9%; specificity 99·0%; AUC 0·77 (0·02); the 90% specificity model identified 40 (21%) of individuals positive for COVID-19 at SO −1; correctly identified 105 (56%) of individuals positive for COVID-19 at SO +4 |

(Table 1 continues on next page)

| | Study design / Population and study setting | Wearable device | COVID-19-positive sample/ Total analysed sample size (n/N) | Race or ethnicity n (%) | Reference standard | Timing compared to SO (days)* | Algorithm or statistical model | Key features in the best performing model | Training set, n (%) | Test, validation (internal or external), or comparison set, n (%) | Key findings for the best performing model |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (Continued from previous page) | | | | | | | | | | | |
| Nestor et al (2021)[32]† | Observational and retrospective (same data collection as used in Shapiro et al[19])/ Ambulatory, opt-in study of device users | Fitbit | 204 /32 198 | NR | Self-reported SARS-CoV-2 test or medically diagnosed influenza | Day of symptomatic infection (SO to symptom end) | Model 1 (wearable only data) was (1a) gradient boosted classifier and (1b) gated recurrent unit-decay; model 2 (survey only data) was gated recurrent unit; model 3 was paired gradient boosted classifier and gated recurrent unit | Model 1 included 48 features based on HR, steps, and sleep data; model 2 included survey data (daily symptom history and demographic covariates); model 3 included 48 features and survey data | 11 269 (35%); 35:7·5:7·5:50 split performed five times | 16 099 (50%) in test set one (prospective); 2415 (7·5%) in test set two (retrospective, held-out set); 2415 (7·5%) in validation set (retrospective) | Model 3 sensitivity 0·65 (95% CI 0·19–0·87), specificity 0·69 (95% CI 0·41–0·97); model 3 detects 63·5% of COVID-19-positive cases at SO (*vs* 47·7% for non-COVID-19 positive influenza like illnesses) |
| Quer et al (2021)[34] | Observational and prospective/ Ambulatory, opt-in study of American smart device users | Device-agnostic | 54 /333 | NR | Self-reported COVID-19 test result | NA | Binary classifier | Resting HR, age, sex, cough, fatigue, decreased taste or smell, sleep duration (min), and total step count | 333 (100%) | NA | AUC 0·80 (95% CI 0·73–0·86); sensitivity 0·72 (95% CI 0·59–0·83); specificity 0·73 (95% CI 0·68–0·78); positive predictive value 0·35 (95% CI 0·29–0·41); negative predictive value 0·93 (95% CI 0·90–0·96) |
| Shapiro et al (2021)[19] | Observational and retrospective digital cohort/ Ambulatory, opt-in study of device users | Fitbit | 41/1352 | No American Indian or Alaskan Native, 4 Asian or Pacific Islander (9·8%), 3 Black or African American (7·3%), 4 Hispanic or Latino (9·8%), 3 preferred not to answer (7·3%), 4 unavailable (9·8%), 23 White (56·1%) | Self-reported COVID-19 diagnosis by a health-care practitioner | SO –2 to SO +2 | Multilevel model | Resting HR, week of flu season, day of the week, average activity level in participant's physical state, and participant's baseline activity level | NA | NA | Increased HR in COVID-19-positive cohort; increased sleep persisted for longer in COVID-19-positive cohort; COVID-19-positive cohort took fewer steps |

(Table 1 continues on next page)

| Study | Study design / Population and study setting | Wearable device | COVID-19-positive sample/Total analysed sample size (n/N) | Race or ethnicity n (%) | Reference standard | Timing compared to SO (days)* | Algorithm or statistical model | Key features in the best performing model | Training set, n (%) | Test, validation (internal or external), or comparison set, n (%) | Key findings for the best performing model |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (Continued from previous page) | | | | | | | | | | | |
| Smarr et al (2020)[38] ‡ | Observational and retrospective/ Global ambulatory, opt-in study of device users | Oura ring | 50/50 | 1 Asian (2%), 39 White (78%), 8 Hispanic or Latino (16%), 1 Middle Eastern (2%),1 European (2%), 1 Scandinavian (2%), 1 Jewish (2%), and 1 South Asian (2%), 2 unavailable (4%) | Self-reported COVID-19 diagnosis or test | SO to SO +7 | Wilcoxon rank-sum test; Kruskal-Wallis non-parametric comparison | Separate models for temperature, respiratory rate, HR, and HRV | NA | NA | Temperature increases around SO; respiratory rate increases after fever-based SO; HR increases after fever-based SO; HRV increases after fever-based SO |

AUC=area under the curve. CuSum=cumulative summary of deviations in elevated residual resting HR. HR=heart rate. HROS-AD=HR over steps anomaly detection. HRV=HR variability. MESOR=midline statistic of rhythm. NA=not applicable. ND=not determined. NR=not reported. RHR-Diff=resting HR difference. RMSSD=root mean square of successive differences in normal heartbeats. SO=symptom onset. $SpO_2$=oxygen saturation. *SARS-CoV-2 infection detection timing relative to SO in days (eg, SO –1 indicates 1 day before SO) across all study models. †Preprint. ‡Proof-of-concept study. §Data are for the COVID-19-positive sub-cohort. ¶Indicates differs by analysis.

*Table 1:* **Eligible studies using wearable devices to detect changes in physiological parameters among COVID-19-positive individuals**

self-reported retrospective SARS-CoV-2 infection; none of the studies tested participants for the presence of SARS-CoV-2 antibodies to detect mild or asymptomatic infections for which the participant had not sought diagnostic testing. Researchers commonly used historical information from long-term wearable use to examine changes in physiological parameters in the days before and after a patient's diagnosis or symptom onset. The studies recruited predominantly from European and North American countries. Nine studies examined SARS-CoV-2 infection among the general public, whereas three enrolled health-care professionals.[31,36,37] Three research teams characterised their studies as proof-of-concept studies.[38–40] Four studies were pre-prints.[31,32,36,40]

The participant sample size (n=29 to 32 198), sex ratio (17–70% male and 30–81% female), and mean ages (29–57 years) varied widely between studies. Information on ethnicity and race was collected and analysed in five studies,[19,32,33,37,38] with only two studies recruiting a relatively diverse population.[19,37]

Various wearable devices were investigated across the 12 studies, with bracelet design constituting the most common style. Five studies examined physiological parameter changes exclusively[19,32,33,35,40] or almost exclusively (99%)[31] measured by Fitbit devices. Other, less commonly investigated wrist-worn devices included the WHOOP strap,[30] the Apple watch,[37] and the Empatica E4[36] (one study each). One study examined a smart ring, the Oura,[38] whereas another study analysed data from an unnamed device worn on the user's throat.[39] The final study remained device-agnostic; most participants wore Fitbits (78·4%), but any device that paired with Apple HealthKit or Google Fit met eligibility criteria and was included.[34] The 12 studies examined wearable device-measured physiological changes in respiratory rate,[30,35,38,39] heart rate,[19,31–35,38–40] heart rate variability,[35,37–39] skin temperature,[36,38] and movement[19,31–34,39] (table 2; appendix pp 15–20).

Across studies, the research teams drew on diverse methods for examining wearable devices' ability to detect SARS-CoV-2 infection (appendix pp 10–11). Nine studies used machine learning algorithms to identify how physiological data (supplemented by symptom reports in three studies)[32,34,36] could detect SARS-CoV-2 infection,[30–36,39,40] including an anomaly detection auto-encoder,[40] gradient-boosted classifiers,[30,32] and deep,[36] convolutional,[35] or gated recurrent-unit[32] neural networks. The remaining three studies used statistical analyses, such as mixed-effect models[19,37] and Wilcoxon rank-sum tests.[38] Although some studies examined differences between SARS-CoV-2 infection and other influenza-like illnesses,[19,32,33,40] most authors focused solely on SARS-CoV-2 infection. Nine studies built models to directly compare wearable data from patients positive for SARS-CoV-2 with healthy[33,36,37,39,40] or SARS-CoV-2-negative controls.[19,31,32,34] Eight studies considered intra-participant changes in baseline parameters as they progressed from uninfected to presymptomatic to symptomatic infection.[19,30,32,33,35,37,38,40]

In general, algorithmic models for detecting SARS-CoV-2 infection were developed retrospectively across the nine studies and focused predominantly on symptomatic disease. Except for Quer and colleagues[34] and Cleary and colleagues,[31] each research team employed cross-validation to test their algorithm's generalisability. Four studies randomly split their data into training and validation sets,[32,35,36,40] whereas other researchers tested their algorithm on healthy and COVID-19-negative controls,[33] recruited an independent set of participants,[30] or used a leave-one-out cross-validation.[39] Acknowledging the effects of seasonal and temporal variance on infection models, Nestor and colleagues[32] validated their model on both a retrospective and prospective test set, determined by its chronological order compared with the training and the validation sets. Reflecting the breadth of model specifications, overall accuracy varied greatly across studies (AUCs ranged from 0·52 to 0·92).[34,39] Among articles reporting sensitivity and specificity, the authors seemingly prioritised specificity over sensitivity (figure 1) meaning that with one exception,[36] studies with very high specificity did not achieve comparably high sensitivity. With more input features, models improved in performance. Quer and colleagues[34] showed that although the model ingesting only symptoms (AUC 0·71) performed similar to the model ingesting only wearable

| | Models included in analysis | Device sensors | Manufacturer | Regulatory status | Principle of operation |
|---|---|---|---|---|---|
| Apple Watch[31,34,37] | Unspecified; Apple Watch Series 4 or 5 | Accelerometer, electrical heart sensor,* gyroscope, and photo-plethysmography | Apple | The EU granted European conformity (CE; also known as Conformité Européenne) marking in March, 2019, for ECG app and irregular HR notifications; US FDA approved ECG app for software as a medical device, temporary approval expanded to encompass remote monitoring of heart health during the COVID-19 pandemic | The Apple Watch provides wearers with a wrist-based notification system, transmitting messages and alerts from their smartphone in real-time; it can be worn during physical activity; its battery life ranges from 1·5–18 h; in addition to supporting third-party apps, the Apple Watch includes health-focused proprietary apps; newer models (eg, the Series 6) include blood oxygen and ECG apps, in addition to the widespread irregular heart rhythm alerts |
| E4 wristband[36] | Unspecified | Accelerometer, electrodermal activity and galvanic skin response, event mark button, infrared thermopile, internal clock, and photo-plethysmography | Empatica | The EU granted CE marking to the E4 wristband, in conjunction with the complementary Aura system, in March, 2021, as a class IIa medical device intended to detect and alert users to an early respiratory infection; approval not granted yet by FDA | Lacking a hardware display, the E4 wristband enables the user to record 32 h of continuous data between device charges; it collects data through multiple sensors and transmits them to a cloud platform, storing up to 60 h of data between transfers; the device allows researchers to record biometric data of participants who are wearing the device at home or in the lab and develop their own customised apps to access participant data in real-time |
| Fitbit smartwatches and trackers[19,31–35,40] | Ionic; Charge 3 and Charge 4; Inspire 2 and Inspire HR; Sense; Versa 2 and Versa 3; unspecified | Accelerometer, altimeter,* barometer,* electrical heart sensors,* GPS,* gyroscope,* orientation,* optical HR,* PurePulse 2.0 HR,* $SpO_2$,* and skin temperature* | Fitbit | Approval not granted yet by EU or FDA | All wrist-worn Fitbit devices rely on wearable sensors to track HR, step count, and sleep stage and quality; newer smartwatch versions (eg, Sense and Versa models) also track skin temperature, $SpO_2$ concentrations, and document potential atrial fibrillation episodes; depending on the model, Fitbit displays provide real-time measurement updates related to the wearer's physical activity and smartphone activity; Fitbit devices can be used continuously and paired with a complementary mobile app, lasting up to 6 days between charges |
| Oura Ring[38] | Unspecified | Accelerometer, negative temperature coefficient, photo-plethysmography, and temperature | Oura | Approval not granted yet by EU or FDA | The Oura's finger-worn design emits a physical display; designed for constant wear and is water resistant, the Oura ring has a 5–7 day battery life; the company has created an accompanying mobile app for the Oura ring; users can track their sleep, activity, and so-called readiness scores on their phone; the sleep score reflects how long the user spends in deep, rapid eye movement, and light sleep, in addition to providing personalised tips for maximising rest; the activity score considers the user's daily steps, calories burned, and amount of time spent inactive; finally, the readiness score gives users a numeric estimate from 0 to 100 of how much their body has recovered from previous activity |
| WHOOP Strap[30] | Unspecified | Accelerometer, capacitive touch, gyroscope, photo-plethysmography, and thermometer | WHOOP | Approval not granted yet by EU or FDA | The wrist-worn WHOOP Strap collects physiological data continuously through multiple sensors; with no digital display on its hardware, the WHOOP strap's battery lasts 4–5 days; when synced with the complementary smartphone app, the WHOOP system quantifies the user's sleep quality, provides recommendations on how much physical exertion could be tolerated, and measures resting HR and HRV; the WHOOP app also enables users to log specific behaviours in a journal each day |

Only the named wearable devices, based on the relevant included literature, are described in the table; thus, the unnamed throat-worn patch (Lonini et al, 2021)[35] is not presented here. ECG=electrocardiogram. FDA=Food and Drug Administration. HR=heart rate. HRV=heart rate variability. $SpO_2$=oxygen saturation. *Model-dependent sensors.

***Table 2:*** **Summary of the wearable devices discussed by name in the included literature, their sensors, and principles of operation**

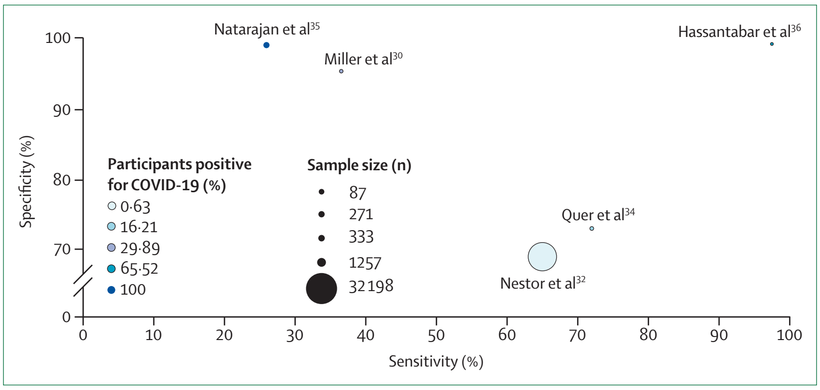


*Figure 1:* **Comparison of the sensitivity and specificity of different machine learning models used for early SARS-CoV-2 detection**

The size of the circle representing each study is proportional to its number of participants. The colour of the circle is proportional to the percentage of participants positive for SARS-CoV-2 in the study.

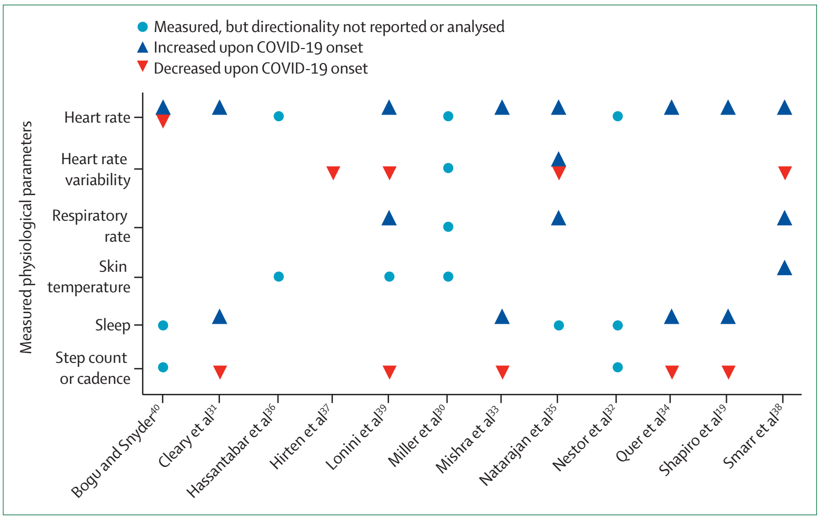


*Figure 2:* **An overview of the main physiological parameters analysed across different studies**

The SARS-CoV-2 associated changes in physiological parameters are shown with upward triangles (indicating a value increase), downward triangles (indicating a value decrease), and circles (indicating parameters were analysed in the study but direction of change was not reported). Notably, Bogu and Snyder's[40] algorithm found bidirectional heart rate abnormalities compared with baseline measurements. Similarly, Natarajan and colleagues[35] report an overall increase in heart rate variability due to COVID-19, despite an initial decrease.

sensor data (AUC 0·72), ingesting both symptoms and sensor data led to superior model performance (AUC 0·80). One cross-sectional study combined data from three separate devices and a self-report questionnaire to achieve an accuracy of 98·1%, compared with 82·4% when relying solely on wearable sensor data.[36] A study enrolling patients with an influenza-like illness episode, which included COVID-19-positive individuals, showed that the symptom-based model (AUC 0·78) outperformed the wearable-based model (sensitivity 0·52, false positive rate 0·4) in distinguishing between COVID-19 cases and non-COVID-19 influenza-like illness cases.[32] With one rare exception,[33] the best performing models (ie, those with >90% specificity[35] and recall of ≥80%[30]) detected a COVID-19 infection 3–7 days after symptom onset.[30,34,35]

The accumulated evidence suggests a trade-off between a model's accuracy and its ability to identify SARS-CoV-2 infection before symptom onset. Only four of the reviewed studies developed models that could detect an impending symptomatic SARS-CoV-2 infection,[33,35,38,40] ranging from 14 days[33] to the day before symptom onset.[35] The algorithms' ability to detect presymptomatic infection also spanned a broad range (20–88% of SARS-CoV-2 infections);[30,33,35,40] however, the greater the number of days preceding symptom onset, the fewer COVID-19 cases a model could identify. For example, Mishra and colleagues[33] detected physiological anomalies in 88% of COVID-19 cases (22 of 25 individuals with a symptom onset date) a median of 4 days (IQR –7 to 0) before symptom onset with their model, whereas Bogu and Snyder[40] reported detecting 56% of COVID-19 cases (14 of 25 individuals) a median of 6·94 days (IQR –7 to –6·22) before symptom onset.

Heart rate, heart rate variability, respiratory rate, skin temperature, and activity levels comprised the most commonly reported physiological parameters measured by wearable devices (figure 2). We discuss the three physiological metrics that could serve as leading indicators of a SARS-CoV-2 infection, and other parameters are reviewed in the appendix (pp 15–20).

Eight articles examining data from more than three wearable devices collectively showed a positive association between SARS-CoV-2 infection and elevated heart rate.[19,31,33–35,38–40] Smarr and colleagues[38] calculated baseline physiological measurements for each Oura-wearing participant (n=50), comparing them to their mean heart rate during the first week of symptomatic infection. They found no significant difference in heart rate during illness based on participants' self-reported symptom onset date (p=0·13), but an association with an increase in heart rate when paired with the start of device-measured temperature shifts (p=0·02). Mishra and colleagues[33] integrated heart rate and step data from 32 Fitbit users to generate a novel heart rate over steps feature. Their analysis revealed that, among 25 individuals with discernible changes in their physiological parameters around symptom onset, heart rate increased by a median of 7 beats per min. Using a subset of Mishra and colleagues' data,[33] Bogu and Snyder[40] developed an algorithm to detect anomalies in resting heart rate around the time of a potential SARS-CoV-2 infection and reported that COVID-19-positive individuals had more recorded hours of abnormal heart rate during the infectious period than healthy peers or those who were ill from a cause other than COVID-19.

Although heart rate anomalies could help alert a wearable device user to an impending infection, research suggests changes in heart rate alone cannot differentiate a SARS-CoV-2 infection from other influenza-like illnesses. Shapiro and colleagues[19] showed that both patients with COVID-19 and patients with influenza had elevated heart rate following self-reported symptom onset. In their device-agnostic studies, Quer and colleagues[34] and Cleary and colleagues[31] found no relative difference between elevated heart rate in COVID-19-negative cohorts and COVID-19-positive cohorts (p=0·33 and p=0·18).[31,34] Furthermore, the same machine learning model ingesting a heart-rate-derived feature could not discriminate well between COVID-19-positive individuals and COVID-19-negative individuals (AUC 0·52 and 0·63).[31,34] In another study,[39] even variability in heart rate before and after activity remained similar, regardless of health status. Converging evidence suggests intrapersonal heart rate might increase following a SARS-CoV-2 infection, but it cannot serve as the sole discriminating factor.

Three of the four studies examining SARS-CoV-2 infection's effect on respiratory rate found that it increased around symptom onset.[35,38,39] In one study,[38] SARS-CoV-2-positive Oura users had higher respiratory rates during the early symptomatic period than during the pre-illness baseline (p=0·002). Training a convolutional neural network on physiological data from 1257 Fitbit wearers, Natarajan and colleagues[35] reported that, during a SARS-CoV-2 infection, respiratory rate deviated from its baseline value more than other parameters. In contrast, Miller and colleagues[30] did not identify respiratory rate as a leading indicator of a potential SARS-CoV-2 infection in their examination of 271 WHOOP strap users who reported COVID-19 symptoms; compared with other physiological parameters, respiratory rate had the lowest coefficient of intraindividual variance over time, regardless of whether the patient was healthy or ill on a given day.

Whereas other articles considered deviations in respiratory rate during a SARS-CoV-2 infection compared with a previous baseline period, Lonini and colleagues[39] examined physiological changes occurring on the same day before and after a given activity. The researchers equipped 15 participants with SARS-CoV-2 infection and 14 healthy participants with an unnamed wearable device. Patients positive for SARS-CoV-2 had similar respiratory rate variability in response to exercise compared with healthy peers (p=0·095), despite a higher baseline value. Cohort demographic differences, however, limit the generalisability of their findings, as most COVID-19 cases had a comorbidity that could have affected their baseline respiratory rate (eg, asthma).[39]

Although fever was one of the first COVID-19 symptoms identified by WHO,[24] of the studies that measured skin temperature, only Smarr and colleagues[38] focused on assessing deviations in this physiological parameter. They compared Oura users' baseline skin temperature to the period following self-reported COVID-19 symptom onset. Statistical analysis revealed an increase in temperature during a symptomatic SARS-CoV-2 infection (p=0·024), with 76% (38 of 50) of participants registering an increase in temperature in the days preceding symptom onset.

We evaluated risk of bias on the basis of the National Institutes of Health's Quality Assessment Tool for Observational Cohort and Cross-sectional Studies.[29] We provide a study-by-study breakdown and detailed descriptions of individual biases in the appendix (pp 21–24). In general, most studies presented a moderate risk of bias; the definition, size, self-reporting of diagnosis, and demographics of the study populations represented a major source of potential bias. Several articles did not clearly define the study population (eg, age, comorbidities, and nationality).[19,30,32,33,40] Three studies also had small samples (total analysed sample n=<1500), despite starting from very large recruited populations (>30 000 individuals).[19,34,35] Some researchers attempted to address the restricted sample size and class imbalance (ie, the number of participants who were positive and negative for COVID-19, or positive and negative days, depending on the type of observation that was analysed) in their algorithms by upsampling infection days,[30] implementing bootstrapping with replacement,[39] or generating a synthetic training dataset.[36] Moreover, most studies identified SARS-CoV-2 infection through participant self-report,[19,30–35,37,38,40] which is overly reliant on subjective data and potentially misses asymptomatic cases. Confounding represented a source of bias faced by many studies, given their restricted adjustment for major demographic factors. Furthermore, many pre-existing comorbidities (eg, body-mass index)[53] shown to affect COVID-19 vulnerability and severity[54] were rarely ingested by the algorithms.[31,35,36]

In addition to the articles detailing completed research, 12 study protocols met inclusion criteria.[41–52] The protocols are investigating numerous wearable devices, ranging from a repurposed fertility tracking bracelet[43,49] to a wearable device supplemented by a sensor placed under the participant's mattress.[45] These studies aimed to assess changes in physiological parameters commonly examined by the other studies we included in our analysis, including heart rate, heart rate variability, and temperature. At least one protocol intends to examine a previously unreported parameter (ie, blood pulse wave).[50] Unlike the completed studies, all protocols propose prospective studies, including two RCTs.[43,50] Two protocols plan to include healthy control groups.[46,51]

## Discussion

This systematic review examined 12 publications and preprints and 12 study protocols related to wearable devices' ability to detect a potential SARS-CoV-2 infection. We observed large variability in device type, physiological parameters analysed, and the operationalisation of diagnostic accuracy across models. Some authors relied on statistical analysis to detect differences between or within participants, whereas others used machine learning algorithms. Accordingly, models varied in their feature specification and performance.

At present, the overall body of evidence regarding the use of wearable devices to detect COVID-19 shows promising, albeit early stage, findings. Most studies drew on retrospective data, had small sample sizes, and did not examine physiological differences from other influenza-like illnesses. Although some studies used PCR testing to confirm SARS-CoV-2 infection, this practice was not universally deployed, potentially introducing diagnostic biases and restricting the comparability of studies to each other. Only three of the included studies explicitly reported using PCR testing—either a weekly PCR test,[37] a per occurrence self-reported PCR test,[35] or a one-time PCR test upon hospital arrival.[36] The fact that two of those studies were conducted solely[37] or partly[36] on health-care professionals shows that this type of population might have had easier access to PCR testing, because of job requirements during the first wave of the COVID-19 pandemic, than the general population. Each study had a different design (ie, prospective,[37] retrospective,[35] and cross-sectional)[36] and investigated the ability of different devices (ie, Apple-Watch,[37] Fitbit,[35] and Empatica E4[36]) to detect deviations in physiological parameters associated with a SARS-CoV-2 infection. Of note, the only included studies that present findings for participants who are infected but asymptomatic, report the use of PCR testing as the reference test.[36,37] Their relatively small sample sizes and prospective and cross-sectional designs could have made it feasible to require PCR testing during data collection. Also, two studies that used PCR tests to determine infection and that aimed to classify the current infection status of participants by developing neural networks, achieved high accuracy (98·1%[36] and 77·0%[35]) and specificity (99·0%).[35,36] However, this performance cannot necessarily be attributed to reliance on the gold-standard PCR tests as an infection marker, as multiple other differences in the specifications and inputs to their models could have influenced their capabilities for detecting a SARS-CoV-2 infection. For example, Natarajan and colleagues[35] enrolled a large sample (n=1257) of symptomatic Fitbit users and examined the classification of a given day for each individual as healthy or ill based on preceding physiological data of heart rate, heart rate variability, and respiratory rate, as well as demographic characteristics. In contrast, Hassantabar and colleagues[36] analysed a much smaller sample (n=87) of healthy (negative PCR test) and symptomatic and asymptomatic patients infected with SARS-CoV-2 (positive PCR test) by monitoring data for up to 1 h from multiple devices measuring galvanic skin response, $SpO_2$, and blood pressure, and collecting questionnaire data on demographics, symptoms, and comorbidities. Despite disparate methods and differences in follow-up time, both studies validated high-performing machine learning algorithms for diagnosing a SARS-CoV-2 infection. Less than half of the studies included a control group of healthy participants, which further limits generalisability.[33,36,37,39,40] Findings from several studies on changes in physiological parameters might also appear contradictory at first glance. However, discrepancies in the direction or magnitude of change could be attributable to the brand or model of a given device. For example, both Miller and colleagues[30] and Natarajan and colleagues[35] analysed data from a wrist-worn device and arrived at differing conclusions about the effect of SARS-CoV-2 infection on respiratory rate. However, the sensors in the specific hardware or the underlying data extraction techniques for interpreting raw sensor data could vary substantially between the WHOOP strap studied by Miller and colleagues[30] and the Fitbit bracelet studied by Natarajan and colleagues.[35] Both wearable devices have a photoplethysmography sensor; however, differences in sampling rates for the sensors could explain variations in interbeat intervals and derived respiratory rates. Device-agnostic studies, pooling data from multiple device models and brands, might seem to directly address these discrepancies through uniform data processing and algorithm development; yet similar concerns could nevertheless render their results difficult to interpret. Recruiting Fitbit users and collecting data from participants' Apple HealthKits and Google-based devices, Quer and colleagues[34] did not correct for potential confounding biases related to the different wearables. Their finding of no changes in heart rate on the basis of SARS-CoV-2 infection status could have derived from how each wearable device measures and processes its raw physiological data. Subsequent device-agnostic studies could further clarify the relationship between seemingly discrepant findings by conducting a head-to-head comparison and determining whether a model's performance varied by device type.

Beyond biases introduced by differences in the studied wearable devices and their associated sensors, the included articles also lacked standardisation in their algorithm development and reporting of performance metrics. After receiving the raw physiological data from a wearable device, researchers make decisions regarding the signal's preprocessing and cleaning (eg, normalising[30,34] or transforming data before model training).[32] Additionally, researchers must choose which optimiser to use when training their model; different optimiser selections can affect model fit and performance. The included studies used more than five types of optimisers in training their respective models. Best practice in machine learning also suggests having a separate test or validation dataset from the training dataset; although two studies did not include any test set,[31,34] seven articles varied greatly in their approach to validating their machine learning algorithm,[30,32,33,35,36,39,40] and the remaining three studies used statistical analyses rather than machine learning methods.[19,37,38] Miller and colleagues[30] tested their model on a dataset of participants derived from the same population as their training data, although their data were recorded during a different time period; in contrast, Hassantabar and colleagues[36] relied on a categorically different population (ie, healthy controls) to validate their algorithm. How each research team chose their test or validation sets inherently influenced their algorithm's performance. Showing how the algorithm development process affects performance, Nestor and colleagues[32] reported their model's sensitivity, specificity, and other metrics if held to the same evaluation schemes as other authors.[30,33,35] Their best performing model achieved higher sensitivity and specificity compared with Natarajan and colleagues,[35] and higher sensitivity and similar specificity compared with Miller and colleagues.[30] Recently, the scientific community has recognised the need for improved standardisation in algorithm development and performance metric reporting, particularly as they relate to health outcomes; multiple publications have called for clinical trials to use machine learning techniques to show their data input process, handling of missing or poor-quality data, and outcomes.[55,56] Although aligned with current practices in machine learning, the varied approaches to algorithm development documented across the included studies could have introduced bias in the findings.

Except for two papers,[36,37] the examined studies also did not consider how physiological parameters might differ between symptomatic and asymptomatic SARS-CoV-2 infections. Most authors trained their algorithms exclusively on symptomatic infections. Identifying asymptomatic infections and building a corresponding model requires testing participants repeatedly for SARS-CoV-2 antibodies; a procedure not done by any of the studies included in our systematic review. Using specialised anomaly-detection algorithms, researchers could train a machine learning model to recognise intrapersonal deviations in physiological parameters during the period between baseline seronegativity and known seroconversion. This model could retrospectively identify the timing of a previous SARS-CoV-2 infection and be applied prospectively to determine real-time asymptomatic—but nevertheless still transmissible—infections. A single protocol, identified by our literature review, has proposed prospective testing and subsequent development of an asymptomatic infection detection algorithm, although it does not specify the method for doing so.[43] Our systematic review suggests wearable devices could help identify SARS-CoV-2 illness before symptom onset, with little self-reported data,[30,33,35,38,40] suggesting their possible usefulness in detecting asymptomatic infection.[37]

An additional challenge identified by this systematic review was that none of the models detecting SARS-CoV-2 infection on the basis of physiological parameters were tested or validated in real-time; although one study tested their online algorithm retrospectively.[33] An algorithm-informed real-time indicator that ingests wearable sensor data could enable individuals to make behavioural changes, such as seeking a SARS-CoV-2 test early and self-isolating. Another study simulated real-world deployment, warning that the shifting prevalence of COVID-19 could cause substantial overestimation of model performance.[32] All identified protocols, however, follow a prospective design, with three protocols aiming to assess an algorithm-driven alert system for health-care professionals[45,50] or the wearable device users.[43,45] Future research plans thus show the need and desire to address this gap.

This systematic review also highlights the disproportionate representation of wrist-worn devices in research surrounding SARS-CoV-2 detection, restricting the results' potential generalisability. Four of the five named devices were smartwatches or wrist-worn straps, whereas only two of the 12 included articles (17%) studied physiological changes related to a SARS-CoV-2 infection with other types of wearable devices (eg, a smart ring).[38,39] We designed our search strategy to minimise potential bias in method of wearable device by including generic terms (eg, remote sensing technology) and searching specifically for non-wrist-worn devices (eg, skin patch and smart glasses; see appendix pp 3–5 for a full list of search terms). Nevertheless, the literature was skewed heavily towards wrist-based wearable devices. An inherent limitation resulting from this disproportionate representation, differences in sensor types, size, and placement could lead to variations in their measurements and accuracy. Although it is beyond the scope of this systematic review to dissect the engineering and design principles varying across wearable devices, we acknowledge the preponderance of wrist-based wearable devices in the summarised literature might unduly influence our findings and conclusions. More studies focused on non-wrist-worn devices will be needed to disentangle how device type influences algorithm performance and which physiological parameters change in relation to SARS-CoV-2 infection. Five protocols identified by our search include non-wrist-worn sensor components, and the results of these studies will contribute much-needed data to this body of evidence.[41,45,46,50,51]

Despite constituting key features affecting participant compliance and overall adoptability, the wearability and perceived ease of use for each studied wearable device were not discussed in any of the included studies. Additionally, comparing usability across devices is difficult because of differences in study design. For example, snapshot studies[36,39] recruiting small samples for a short period of time placed a small burden on participants in terms of time and effort; a participant might be more likely to tolerate an uncomfortable device if they need only to wear it for a few hours compared with a study lasting several months. Conversely, studies using stand-alone consumer wearables relied on users to opt-in to their clinical trials;[19,30,32–35,38,40] participants might have felt more comfortable with the device they already own compared to study participants who were given the device as study material.[31,37] Many of the included studies required a minimum time or days of use of the wearable device as a prerequisite for a participant to be included in the analysed sample.[19,30,35,38] However, there were differences in how long users had to wear the device each day between studies with extended timelines. For example, some studies[30,35] developed their models using only night data, although the devices were designed to be worn throughout the day. Consequently, their findings suggest individuals need only wear the device while at rest to detect a SARS-CoV-2 infection. Usability and perceived ease-of-use would probably be affected by how often a user has to wear the device, in addition to its baseline comfort. Finally, some researchers observed that participants occasionally did not use the wearable device when symptomatic,[33] indicating participant's health could interact with the device's overall wearability, affect data collection, and subsequently the underlying model's ability to detect an infection. This research team also argued that devices requiring daily charging are expected to have more missing data, showing that this feature potentially affects the data quality of the study.[33] Although another research team analysed all participants in the cohort and used machine learning models that can handle implicitly missing data for this purpose, they report a drop in performance if an individual had not worn the device for a week.[32] The fact that several of the included study protocols aim to assess the feasibility of wearing a device for a specific amount of time is encouraging.[41,44,45] Various future studies intend to establish individuals' comfort in following the necessary compliance schedule to maximise the usefulness of a SARS-COV-2 detection algorithm. One trial designed for the specific intensive settings of a 14-day quarantine will instruct participants to wear the device all the time except for when showering and charging the device.[50] Similarly, another planned study will ask participants to wear the device all the time outside of work (health-care workers) for 30 days, but recognises in their protocol that there is a small chance of discomfort or skin abrasion from the prolonged use of the wristband without following appropriate hygiene practices.[52] One trial will aim to balance the prolonged follow-up of the participants with asking them to wear the monitoring bracelet only at night, so that the skin can breathe and dry during the day.[43] Although we cannot comment on the wearability or ease-of-use of the studied devices, future synthesis of these factors for each device should be feasible given the clinical trials underway.

Finally, sample selection and participant demographics limit the models' generalisability across populations. For example, most studies, which were done solely or largely in the USA, had little racial diversity, despite COVID-19 disproportionately affecting Black and Hispanic communities in the USA.[57] Wearable devices have previously shown variable performance across differing skin tones; consequently, if research does not explicitly include these populations, diagnostic accuracy and potential for public good within vulnerable communities remains limited.[58] Similarly, although some models attempted to take into account sex-based variance,[34,35] none of the machine learning algorithms considered how physiological parameters change across the menstrual cycle.[59–61] Future researchers should consider and fine-tune their algorithms to adjust for sex-based differences, thereby reducing the likelihood that a postovulatory shift in temperature, for example, would be erroneously labelled as COVID-19.

Despite these limitations, the reviewed studies provide valuable insights for future research on common wearable-measured physiological parameters. Eight articles showed an increase in heart rate associated with a SARS-CoV-2 infection,[19,32–35,38–40] in line with population-level heart rate data associated with influenza.[62] Similarly, changes in skin temperature and activity frequency provide encouraging results for emerging wearable devices equipped with a temperature sensor and an accelerometer. Notably, behavioural changes, such as receiving a SARS-CoV-2 test result after symptom onset could result in an overestimation of the performance of models based on activity frequency, limiting the use of these data samples.[31] In addition, establishing a conclusive COVID-19-related pattern in respiratory rate and heart rate variability requires additional replications to disentangle contradicting or inconclusive initial findings. Other features, such as coughing patterns from mechanoacoustic sensors could be used to decipher further trends relevant to a SARS-CoV-2 infection,[39] although this possibility remains to be shown. Several experimental papers also used this approach,[63] although their content did not fit the inclusion criteria for this systematic review.

This study represents a comprehensive search of multiple databases and literature to date; we included multiple synonyms for each primary term, tailored the search terms to each database, manually screened reference lists, and actively tried to mitigate any missing data. Owing to the rapid pace of COVID-19 research, we also sought out preprint sources. Despite these strengths, although we did not restrict language in the databases we used, we did not search databases publishing only non-English publications; thus we might have overlooked relevant literature published in another language. Additionally, we identified only studies captured according to our search terms, in specific medical and research databases. Despite our efforts, we might have missed some relevant studies which have not yet been published in peer-reviewed journals or as preprints (eg, findings reported in news articles or company press releases). We sought to mitigate this potential limitation by including a supplementary search for study protocols, which could highlight ongoing clinical trials and potential future publications relevant to our research question.

## Conclusion

Adequately containing the COVID-19 pandemic requires rapid identification of individuals who are infectious. Although wearable devices could help, this systematic review highlights the need for well designed and controlled studies to robustly identify if wearables can accurately detect SARS-CoV-2 infection before symptom onset or in asymptomatic individuals in comparison to the current gold-standard diagnostic method. Future studies should additionally consider how inherent differences in wearable sensor methods, raw data processing, and algorithm development contribute to the detection of infection-associated deviations in physiological measurements and how to address sources of bias.

**Contributors**

All authors had full access to the data, including statistical reports and tables, in the study. MM, BMG, and GSD take responsibility for the integrity of the data and the accuracy of the data analysis. MM, BMG, DEG, PS, MC, and GSD contributed to the study concept and design. MM, BMG, AS, VK, and GSD contributed to the acquisition, analysis, and interpretation of data and provided administrative, technical, and material support. MM, BMG, and GSD contributed to drafting the manuscript. All authors contributed to a critical revision of the manuscript for important intellectual content. DEG, PS, and MC obtained funding. GSD was responsible for study supervision.

**Declaration of interests**

MM, BMG, VK, BF, DV, DEG, MC, and GSD received grants from Innovative Medicines Initiative 2 Joint Undertaking (number 101005177), during the conduct of the study. BMG reports consulting fees and employment from Ava Science, support for attending meetings and travel from Ava Aktiengesellschaft (AG), a patent application from Ava AG (P24892CH00) filed with the Swiss Federal Institute of Intellectual Property for System and Method for Pre-Symptomatic and/or Asymptomatic Detection of a Human Viral or Bacterial Infection based on pilot data from the COVID-RED clinical study, and consultancy for Falcon Health and TheraB Medical, outside the submitted work. VK reports employment from Ava Science and Ava AG, during the conduct of this study. TBB, BF, DV, and DEG report employment from Julius Clinical Research, during the conduct of the study. MC reports employment from Ava AG during the conduct of the study. GSD reports a grant from Health Holland, outside the submitted work. All other authors declare no competing interests.

**Data sharing**

Template data collection forms and the completed data extraction table are included in the appendix (pp 26–40).

**Acknowledgments**

This project was funded by the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement number 101005177. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations. The funding organisation played no part in the design or conduct of the study; collection, analysis, or interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication. The researchers confirm their independence from the funders. We thank Ariel Dowling and Johannes B Reitsma for their comments on an early draft of the Review.

**References**

1 WHO. Novel coronavirus (2019-nCoV) situation report-1. Jan 21, 2020. https://apps.who.int/iris/bitstream/handle/10665/330760/nCoVsitrep21Jan2020-eng.pdf?sequence=3&isAllowed=y (accessed Dec 3, 2021).

2 WHO. WHO coronavirus (COVID-19) dashboard. 2020. https://covid19.who.int/ (accessed March 16, 2022).

3 Kucharski AJ, Klepac P, Conlan AJK, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis* 2020; **20:** 1151–60.

4 Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *Lancet Public Health* 2020; **5:** e452–59.

5 WHO. Laboratory testing for coronavirus disease (COVID-19) in suspected human cases: interim guidance. March 19, 2020. https://apps.who.int/iris/handle/10665/331501 (accessed March 16, 2022).

6 Corman VM, Landt O, Kaiser M, et al. Detection of 2019 -nCoV by RT-PCR. *Euro Surveill* 2020; **25:** 1–8.

7 US Food and Drug Administration. Coronavirus testing basics. May, 2020. https://www.fda.gov/media/138094/download (accessed Dec 3 2021).

8 Dinnes J, Deeks JJ, Adriano A, et al. Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev* 2020; **8:** CD013705.

9 Elias C, Sekri A, Leblanc P, Cucherat M, Vanhems P. The incubation period of COVID-19: a meta-analysis. *Int J Infect Dis* 2021; **104:** 708–10.

10 Walsh KA, Jordan K, Clyne B, et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *J Infect* 2020; **81:** 357–71.

11 European Centre for Disease Prevention and Control. COVID-19 testing strategies and objectives. Sept 15, 2020. https://www.ecdc.europa.eu/sites/default/files/documents/TestingStrategy_Objective-Sept-2020.pdf (accessed 3rd Dec 3 2021).

12 Centers for Disease Control and Prevention. COVID-19 testing: what you need to know. 2021. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html (accessed March 23, 2021).

13 National Institute for Public Health and the Environment. Testing for COVID-19. 2021. https://www.rivm.nl/en/novel-coronavirus-covid-19/testing-for-covid-19 (accessed March 23, 2021).

14 National Health Service. Testing for coronavirus (COVID-19). 2021. https://www.nhs.uk/conditions/coronavirus-covid-19/testing/ (accessed March 23, 2021).

15 Zou L, Ruan F, Huang M, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med* 2020; **382:** 1177–79.

16 He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* 2020; **26:** 672–75.

17 Allen WE, Altae-Tran H, Briggs J, et al. Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat Hum Behav* 2020; **4:** 972–82.

18 Centers for Disease Control and Prevention. Similarities and differences between flu and COVID-19. 2021. https://www.cdc.gov/flu/symptoms/flu-vs-covid19.htm (accessed Dec 3, 2021).

19 Shapiro A, Marinsek N, Clay I, et al. Characterising COVID-19 and influenza illnesses in the real world via person-generated health data. *Patterns (N Y)* 2020; **2:** 100188.

20 Ing AJ, Cocks C, Green JP. COVID-19: in the footsteps of Ernest Shackleton. *Thorax* 2020; **75:** 693–94.

21 Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *J Assoc Med Microbiol Infect Dis Canada* 2020; **5:** 223–34.

22 Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. *PLoS Med* 2020; **17:** e1003346.

23 Johansson MA, Quandelacy TM, Kada S, et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Netw Open* 2021; **4:** e2035057.

24 WHO. Report of the WHO–China joint mission on coronavirus disease 2019 (COVID-19). Feb 28, 2020. https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19) (accessed Dec 3, 2021).

25 Zhu TY, Rothenbühler M, Hamvas G, et al. The accuracy of wrist skin temperature in detecting ovulation compared to basal body temperature: prospective comparative diagnostic accuracy study. *J Med Internet Res* 2021; **23:** e20710.

26 Chen G, Xie J, Dai G, et al. Validity of the use of wrist and forehead temperatures in screening the general population for covid-19: a prospective real-world study. *Iran J Public Health* 2020; **49** (suppl 1)**:** 57–66.

27 Seshadri DR, Davies EV, Harlow ER, et al. Wearable sensors for COVID-19: a call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments. *Front Digit Health* 2020; **2:** 8.

28 Mitratza M, Goodale BM, Downward GS, Stolk P, Shagadatova A. Performance of wearable sensors in the detection of COVID-19: a systematic review. 2021. https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021232910 (accessed March 16, 2022).

29 National Heart, Lung, and Blood Institute. Study quality assessment tool for observational cohort and cross-sectional studies. 2014. https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort (accessed March 23, 2021).

30 Miller DJ, Capodilupo JV, Lastella M, et al. Analyzing changes in respiratory rate to predict the risk of COVID-19 infection. *PLoS One* 2020; **15:** e0243683.

31 Cleary JL, Fang Y, Sen S, Wu Z. A caveat to using wearable sensor data for COVID-19 detection: the role of behavioral change after receipt of test results. *medRxiv* 2021; published online April 22. https://www.medrxiv.org/content/10.1101/2021.04.17.21255513v1 (preprint).

32 Nestor B, Hunter J, Kainkaryam R, et al. Dear watch, should I get a COVID-19 test? Designing deployable machine learning for wearables. *medRxiv* 2021; published online May 17. https://www.medrxiv.org/content/10.1101/2021.05.11.21257052v1 (preprint).

33 Mishra T, Wang M, Metwally AA, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat Biomed Eng* 2020; **4:** 1208–20.

34 Quer G, Radin JM, Gadaleta M, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nat Med* 2021; **27:** 73–77.

35 Natarajan A, Su HW, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. *NPJ Digit Med* 2020; **3:** 156.

36 Hassantabar S, Stefano N, Ghanakota V, et al. CovidDeep: SARS-CoV-2/COVID-19 test based on wearable medical sensors and efficient neural networks. *arXiv* 2020; published online Oct 28. https://arxiv.org/abs/2007.10497 (preprint).

37 Hirten RP, Danieletto M, Tomalin L, et al. Use of physiological data from a wearable device to identify SARS-CoV-2 infection and symptoms and predict COVID-19 diagnosis: observational study. *J Med Internet Res* 2021; **23:** e26107.

38 Smarr BL, Aschbacher K, Fisher SM, et al. Feasibility of continuous fever monitoring using wearable devices. *Sci Rep* 2020; **10:** 21640.

39 Lonini L, Shawen N, Botonis O, et al. Rapid screening of physiological changes associated with COVID-19 using soft-wearables and structured activities: a pilot study. *IEEE J Transl Eng Health Med* 2021; **9:** 4900311.

40 Bogu GK, Snyder MP. Deep learning-based detection of COVID-19 using wearables data. *medRxiv* 2021; published online Jan 9. https://www.medrxiv.org/content/10.1101/2021.01.08.21249474v1 (preprint).

41 Choi SW, Briskin ES, Schlafer SK. Monitoring health care workers at risk for COVID-19 using wearable sensors and smartphone technology. Feb 16, 2021. https://clinicaltrials.gov/ct2/show/NCT04756869 (accessed Dec 3, 2021).

42 Woods C, Dunn J, Shaw R. Using smart watches to detect and monitor COVID-19 (CovIdentify). 2020. https://clinicaltrials.gov/ct2/show/NCT04623047 (accessed July 27, 2021).

43 Brakenhoff TB, Franks B, Goodale BM, et al. A prospective, randomized, single-blinded, crossover trial to investigate the effect of a wearable device in addition to a daily symptom diary for the remote early detection of SARS-CoV-2 infections (COVID-RED): a structured summary of a study protocol for a randomized controlled trial. *Trials* 2021; **22:** 1–5.

44 Cislo C, Clingan C, Gilley K, et al. Monitoring beliefs and physiological measures using wearable sensors and smartphone technology among students at risk of COVID-19: protocol for a mHealth study. *JMIR Res Protoc* 2021; **10:** e29561.

45 Frasch MG. Non-invasive biometric monitoring in nursing homes to fight COVID-19. 2021. https://clinicaltrials.gov/ct2/show/NCT04548895 (accessed July 27, 2021).

46 Jayaraman A. Wearable sensor to monitor COVID-19 like signs and symptoms. 2020. https://clinicaltrials.gov/ct2/show/NCT04393558 (accessed July 27, 2021).

47 Mault J. Wearable diagnostic for detection of COVID-19 infection. 2021. https://clinicaltrials.gov/ct2/show/NCT04742569 (accessed July 27, 2021).

48 Ramirez E, Foschini L. A virtual prospective study exploring activity trackers and COVID-19 infections. 2020. https://clinicaltrials.gov/ct2/show/NCT04623138 (accessed Dec 3, 2021).

49 Risch L, Conen D, Risch M, Aeschbacher S, Grossman K. Can the Ava fertility tracker device detect early signs of COVID-19? 2020. http://www.isrctn.com/ISRCTN51255782 (accessed Dec 3, 2021).

50 Wong CK, Ho DTY, Tam AR, et al. Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial. *BMJ Open* 2020; **10:** e038555.

51 Xu S. Vital sensor monitors for CV19 detection. 2020. https://clinicaltrials.gov/ct2/show/NCT04635787 (accessed Dec 3, 2021).

52 Zargaran A, Radenkovic D, Theodoulou I, et al. The COVID-19 early detection in doctors and healthcare workers (CEDiD) study: study protocol for a prospective observational trial. *medRxiv* 2020; published online Dec 10. https://www.medrxiv.org/content/10.1101/2020.08.11.20172502v2 (preprint).

53 Popkin BM, Du S, Green WD, et al. Individuals with obesity and COVID-19: a global perspective on the epidemiology and biological relationships. *Obes Rev* 2020; **21:** e13128.

54 Zhou Y, Yang Q, Chi J, et al. Comorbidities and the risk of severe or fatal outcomes associated with coronavirus disease 2019: a systematic review and meta-analysis. *Int J Infect Dis* 2020; **99:** 47–56.

55 Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020; **26:** 1364–74.

56 Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* 2020; **26:** 1351–63.

57 Renelus BD, Khoury NC, Chandrasekaran K, et al. Racial disparities in COVID-19 hospitalization and in-hospital mortality at the height of the New York city pandemic. *J Racial Ethn Health Disparities* 2021; **8:** 1161–67.

58 Colvonen PJ, DeYoung PN, Bosompra NA, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. *Sleep* 2020; **43:** 1–3.

59 Goodale BM, Shilaih M, Falco L, Dammeier F, Hamvas G, Leeners B. Wearable sensors reveal menses-driven changes in physiology and enable prediction of the fertile window: observational study. *J Med Internet Res* 2019; **21:** e13404.

60 Shilaih M, Goodale BM, Falco L, Kübler F, De Clerck V, Leeners B. Modern fertility awareness methods: wrist wearables capture the changes of temperature associated with the menstrual cycle. *Biosci Rep* 2018; **38:** BSR20171279.

61 Shilaih M, Clerck V, Falco L, Kübler F, Leeners B. Pulse rate measurement during sleep using wearable sensors, and its correlation with the menstrual cycle phases, a prospective observational study. *Sci Rep* 2017; **7:** 1294.

62 Radin JM, Wineinger NE, Topol EJ, Steinhubl SR. 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* 2020; **2:** e85–93.

63 Ni X, Ouyang W, Jeong H, et al. Automated, multiparametric monitoring of respiratory biomarkers and vital signs in clinical and home settings for COVID-19 patients. *Proc Natl Acad Sci USA* 2021; **118:** 1–12.