

Can Wearables Detect Physiological Deterioration Before Symptoms?

ABSURD Health Institute · Volume Two · Research Essay 55

Research series: 08 · Signals, measurement and responsible medical AI

Abstract

Consumer wearables can gather heart rate, movement, sleep estimates and in some devices temperature-related signals over long periods. Changes relative to baseline may precede recognized illness in selected contexts. This paper examines the benefits and pitfalls of applying anomaly detection to personal physiology and argues that actionable warning depends on specificity, sustained change and clinical context.

A wearable is an early-warning sensor, not a diagnosis

Wearables could identify an evolving physiological disturbance before ordinary symptom awareness, but baseline shifts alone are not disease diagnoses.

Resting heart rate may rise during fever, stress, dehydration, poor sleep or certain medications. Reduced activity can be both a cause and consequence of illness. Sleep staging from consumer devices is an algorithmic estimate, not the equivalent of full polysomnography. Sensor characteristics vary across hardware versions and body placement; missing data are often not random because people remove devices when uncomfortable.

The conceptual opportunity lies in within-person deviations sustained across multiple modalities. A brief rise in heart rate after exercise should be treated differently from an unexplained multi-day shift accompanied by increased temperature and reduced recovery. However, even converging signals rarely identify a unique disease; they primarily indicate that something has changed. The system should separate general wellness alerts from urgent clinical warning signs.

Individual baselines and everyday confounding

Wearable signals can shift in response to illness, but the most difficult part is deciding when a deviation becomes actionable. A nocturnal heart-rate elevation may reflect fever, alcohol, stress, dehydration, travel or poorly fitted sensors. Several changes occurring together may increase concern without identifying the disease. Because a user supplies many days of data, seemingly rare alerts can become common over a year. That creates a longitudinal false-discovery problem unlike a single clinic visit. Designs should incorporate sustained deviation, clinically relevant thresholds and the option for a neutral “watch and reassess” message rather than constant alarm.

Wearables provide continuous or frequent measurements of pulse, movement, sleep and sometimes temperature-related signals, but their sensors are not diagnostic instruments for every disease they might correlate with. Resting pulse can increase during infection, anxiety, dehydration or stimulant

exposure; reduced activity can reflect deliberate rest rather than clinical decline. The baseline period, individual variability and measurement reliability all determine whether a deviation is meaningful. A device that performs well in healthy volunteers may perform differently in arrhythmia or limited mobility.

Prediction must be distinguished from early symptom measurement. If the user already feels ill when the algorithm identifies a pulse change, the alert may be corroborative rather than presymptomatic. A robust evaluation should define a clinically adjudicated endpoint, a lead-time interval and how many alerts were issued per true event. Missing days, uncharged batteries and imperfect wear adherence should be included in performance calculations, not silently discarded.

Heart rate, activity, temperature proxies and sleep estimates change with infection, stress, exercise, travel, medications and sensor placement. A meaningful algorithm must learn individual rhythm while distinguishing actionable deterioration from expected variation. The useful frontier claim concerns early warning lead time and practical response, not the assertion that a watch can diagnose serious disease from one elevated pulse.

Alerts per year versus actionable benefit

An intervention study should establish what the wearer or care team is supposed to do after an alert. Uncertain notifications can increase visits and anxiety without reducing serious outcomes. The central opportunity lies in intelligently combining longitudinal trends with symptom check-ins and individualized thresholds, but improved patient outcomes remain the proof of benefit. Wearables are most credible as supplementary risk signals, not substitutes for examination or emergency triage.

Signal frequency changes statistical burden. A wearable collecting thousands of observations per day has many opportunities to register apparent anomalies. The data are correlated, and ordinary daily activity can create prolonged deviations. An alert algorithm should interpret episodes and context rather than count every unusual sample as independent evidence. Individualized baselines are promising but can drift if someone becomes chronically ill, causing the system to normalize a dangerous trend.

What counts as lead time. The date of clinical diagnosis is often later than the biological onset of illness. An algorithm that flags a change before diagnosis may still be detecting symptoms the wearer noticed but had not reported. Studies should capture self-reported symptom onset and blinded clinical outcomes alongside sensor data.

Proving net benefit. Compare alert-driven care with usual care while preserving clear safety escalation routes. Measure true actionable detections, false alarm burden, missed events, unnecessary investigations and outcomes after intervention. A device that predicts with modest accuracy may be useful if it detects a high-consequence event cheaply, whereas a more accurate model may be impractical if its alerts lead nowhere. Product claims must align with this evidence rather than with the appearance of sophisticated graphs.

Evaluate pre-registered alert thresholds against independently adjudicated illness episodes, include user-specific baseline adaptation and compare to a simple symptom checklist. Record false alarms,

missed events and unnecessary downstream testing. Demonstrating earlier alerts without a benefit from the actions they trigger is an incomplete clinical argument.

Evidence for pre-symptomatic pattern changes is not evidence that automatic alerts improve outcomes. Validation should examine lead time, false-alert frequency per person-year, subgroup performance, adherence and clinical consequences. The safest model is an assistant to observation and medical triage, not a replacement for symptom evaluation or urgent medical care.

An alert can arrive earlier and still make the patient worse off

A wearable may notice a higher resting heart rate before a person reports being ill. But exercise, alcohol, travel, heat, anxiety and device noise can produce similar deviations, especially in data collected continuously. In a year a seemingly modest false-alert rate per day can yield many unnecessary medical encounters. The strongest proposed early-warning system adapts to each user's baseline and looks for a coherent trajectory across multiple reliable measurements. It should specify what action follows an alert and whether that action reduces complications or improves function. A prospective evaluation would compare alert timing, missed events and downstream testing with a simple rule or symptom checklist. Longitudinal sensitivity alone does not justify deployment; false-positive burden and the cost of investigation matter. A credible system can succeed by detecting a narrow, serious phenotype accurately instead of promising universal disease recognition from consumer sensors.

Earlier is only better if useful

A continuous wearable can detect deviations that would be missed during infrequent consultations, but the same frequency increases opportunities for nuisance alerts. A shift in pulse or sleep may reflect real disease, changed activity, stress or sensor error. Building a useful system requires reliable baselines, contextual interpretation and prospective outcome adjudication. Model accuracy should be reported per person and clinical event rather than inflated by counting many correlated samples. A meaningful early-warning claim also requires demonstrating that alerts occur before symptoms or clinical suspicion, not merely before a formal diagnosis code. Trial designs should measure whether alerts lead to appropriate evaluation and better outcomes while recording anxiety, unnecessary testing and missed emergencies. Wearables can complement care through longitudinal context; the proof of value is not that a graph looks medically sophisticated but that it changes decisions for the better.

False alarms can accumulate rapidly in continuously monitored populations, and excessive attention to daily variability may increase health anxiety. A classifier trained to flag infection may also respond to travel, alcohol, medication or menstrual changes. Demonstrating detection before reported symptoms is not equivalent to demonstrating fewer hospitalizations or better outcomes.

This paper provides a safety-aware framework for Apple Watch, Oura and comparable device integrations without implying any specific data connector or product validation. Its strongest insight is that personal baselines can increase signal relevance only when robust uncertainty management and confirmation pathways are built into the system.

Sources and research leads

1. Mishra et al. Pre-symptomatic detection of COVID-19 from smartwatch data
2. Quer et al. Wearable sensor data and self-reported symptoms for COVID-19 detection

Editorial status: Third-pass conceptual research manuscript. The argument has been individually strengthened and shared boilerplate removed. Original reference leads were retained; a complete independent source-by-source scholarly verification has not been performed. No new clinical trial data are presented.