Wearables Detect Malaria Early in a Controlled Human-Infection Study

Sidhartha Chaudhury, Chenggang Yu, Ruifeng Liu ⓘ, Kamal Kumar, Samantha Hornby, Christopher Duplessis, Joel M. Sklar, Judith E. Epstein, and Jaques Reifman ⓘ

Abstract—*Objective:* Observational studies on the use of commercially available wearable devices for infection detection lack the rigor of controlled clinical studies, where time of exposure and onset of infection are exactly known. Towards that end, we carried out a feasibility study using a commercial smartwatch for monitoring heart rate, skin temperature, and body acceleration on subjects as they underwent a controlled human malaria infection (CHMI) challenge. *Methods:* Ten subjects underwent CHMI and were asked to wear the smartwatch for at least 12 hours/day from 2 weeks pre-challenge to 4 weeks post-challenge. Using these data, we developed *2B-Healthy*, a Bayesian-based infection-prediction algorithm that estimates a probability of infection. We also collected data from eight control subjects for 4 weeks to assess the false-positive rate of *2B-Healthy*. *Results:* Nine of 10 CHMI subjects developed parasitemia, with an average time to parasitemia of 12 days. *2B-Healthy* detected infection in seven of nine subjects (78% sensitivity), where in six subjects it detected infection 6 days before parasitemia (on average). In the eight control subjects, we obtained a false-positive rate of 6%/week. *Conclusion:* The *2B-Healthy* algorithm was able to reliably detect infection prior to the onset of symptoms using data collected from a commercial smartwatch in a controlled human infection study. *Significance:* Our findings demonstrate the feasibility of wearables as a screening tool to provide early warning of infection and support further research on the use of the *2B-Healthy* algorithm as the basis for a wearable infection-detection platform.

Index Terms—Detection, infection, malaria, physiological monitoring, wearable device.

I. INTRODUCTION

WEARABLE devices that continuously monitor vital signs can serve as powerful tools for early detection of adverse health events [1]. We have previously shown that wearable devices can be used to detect heat stress [2], [3], blood glucose deregulation [4], and sleep-associated performance deficits [5]. Early detection of infection has the potential to provide many benefits, including automated screening and isolation, leading to a reduction in the spread of infectious diseases, as well as early diagnosis and intervention, leading to improved treatment outcomes. Spurred by the COVID-19 pandemic, there has been a renewed focus on the use of wearable devices to detect infection [6]–[9], with some studies enrolling thousands of subjects to voluntarily provide wearable data collected from commercially available devices, such as the Fitbit, along with symptom surveys and testing results. While these observational studies are vital for demonstrating the potential large-scale use of wearable devices for infection detection, they have several limitations.

First, these studies are based on self-reported symptoms and infection cases, and testing is typically only sought after subjects are symptomatic. While such studies can be useful to determine time-to-detection prior to symptom onset, they do not allow us to determine the time-to-detection from exposure because the time of exposure is not known. Second, detection sensitivity in these studies is biased towards symptomatic cases of infection because asymptomatic cases would typically be untested and thus unreported. Third, assessment of detection specificity in these studies is difficult because of the lack of knowledge of the non-reporting rate—negative subjects who had symptoms but did not report them or were tested but not counted. In contrast to these observational studies, controlled human-infection studies, in which human subjects are deliberately infected with a pathogen in a controlled clinical environment and then closely monitored

for symptoms and infection, provide an essential complementary approach for developing and evaluating infection-detection technologies.

Controlled human malaria infection (CHMI) studies are an ideal vehicle for developing and assessing wearable-based infection-detection technologies. CHMI is a well-established clinical model first developed by the U.S. Army, the U.S. Navy, and the National Institutes of Health in 1986 [10], which provides clearly defined symptomology for a severe infectious disease of clinical and military relevance. In a typical CHMI study, subjects are deliberately exposed to *Plasmodium falciparum* through a range of vehicles, including direct venous inoculation or infected mosquito bites. Following entry into the body, the *Plasmodium* parasite transits via the blood to the liver (liver-stage infection), where it resides for 9-14 days (pre-patent period or incubation stage), before developing into the blood-stage parasite (blood-stage parasitemia), which can be highly lethal if left untreated. From 7 days post-challenge, subjects are tested daily for blood-stage parasitemia and are immediately treated with FDA-approved antimalarial drugs once it is detected, rapidly clearing the infection. In many CHMI studies, during the second week post-challenge, when the blood-stage infection is expected to emerge, subjects are required to spend their nights in a study-selected hotel for clinical observation. CHMI closely reflects the typical sequence of events that occurs in natural malaria infection, and approximately 95-100% of subjects undergoing CHMI achieve blood-stage parasitemia [11].

The clinical characteristics of CHMI have been reported in a number of investigations that combine observations across multiple CHMI clinical studies [12]–[14]. In a review of 17 CHMI studies, Epstein *et al.* identified several common features associated with CHMI [15]. First, symptom onset occurs, on average, slightly ahead of the onset of parasitemia, as detected by blood smear. Approximately 60% of the subjects reported malaria symptoms at least 12 hours prior to detection of parasitemia, and 30% reported symptoms on the same day that parasitemia was detected. The reported symptoms were classified as "mild" to "moderate" in ∼80% of subjects, although 20% reported having "severe" symptoms, based on criteria defined by the FDA [16]. The most common symptoms are fatigue, headache, malaise, chills, and myalgia, lasting from 2 to 4 days. Epstein *et al.* also noted that body temperatures measured at post-challenge day 10 were greater than 37.5 °C for 80% of subjects, greater than 38.0 °C for 60%, and greater than 40 °C for 5% [15]. In another CHMI study, tachycardia was reported in 33% of the cases (3 of 9), although subjects' vital signs were not continuously monitored [17].

In the present study, we used a commercial-off-the-shelf (COTS) Samsung Gear S3 smartwatch to monitor heart rate, skin temperature, and body acceleration data (i.e., activity via a 3-axis accelerometer) from subjects as they went through CHMI as part of a separate malaria-vaccine study [18]. Subjects were asked to wear the device from 2 weeks pre-challenge to 4 weeks post-challenge as they were monitored for signs and symptoms of malaria, encompassing the time to projected parasitemia, clinical expression of disease, and resolution of illness post treatment. Using these data, we developed the *2B-Healthy* algorithm, which

analyzed the heart rate and actigraphy data from the smartwatch to generate a real-time probability of infection. We compared this probability of infection with the primary clinical endpoint of blood-stage parasitemia, as well as a secondary endpoint of malaria symptom onset. The period between challenge and parasitemia, as detected by blood smear, typically referred to as the pre-patent period in malaria research, will be referred to hereafter as the incubation period. We assessed infection-detection accuracy based on the time-to-detection of infection during the incubation period prior to parasitemia, as determined by blood smear; the time-to-detection prior to onset of malaria-related symptoms; and the false-positive rate as determined from data collected from healthy control individuals who did not undergo CHMI.

A number of key aspects of this study distinguish it from prior wearable infection-detection efforts. First, unlike in studies of natural infection, because we are using a controlled-infection model, the exact time of exposure is known and the progression of the disease is well characterized, enabling us to more precisely link abnormal heart rates with the onset of infection. Second, while prior studies typically report associations between wearable data and infection symptomology, ours is the first to use a Bayesian-based approach that accounts for circadian rhythm and activity level to estimate a probability of infection, enabling model customization for individualized real-time prediction. Finally, our algorithm is designed to be a generalizable, non-specific, infection-screening tool–it is device agnostic and could theoretically be used with any wearable device that accurately measures heart rate and activity data for any kind of infection event that generates abnormalities in heart rate patterns.

II. METHODS

A. Wearable Data Collection

1) Ethics: The study protocol for this clinical trial was approved by the Institutional Review Boards (IRBs) at the Walter Reed Army Institute of Research (WRAIR) and the Naval Medical Research Center (NMRC). The study was conducted at the NMRC Clinical Trials Center in accordance with: the principles described in the Nuremberg Code and the Belmont Report; all federal regulations regarding the protection of human participants as described in 32 CFR 219 (The Common Rule) and instructions from the U.S. Department of Defense, the U.S. Department of the Army, the U.S. Department of the Navy, and the Bureau of Medicine and Surgery of the U.S. Navy; and the internal policies for human subject protections and the standards for the responsible conduct of research of the U.S. Army Medical Research and Development Command and the NMRC. WRAIR holds a Federal Wide Assurance from the Office of Human Research Protections under the Department of Health and Human Services as does the NMRC. The NMRC also holds a Department of Defense/Department of the Navy Federal Wide Assurance for human subject protections. All key personnel were certified as having completed mandatory human research ethics education curricula and training under the direction of the WRAIR IRB or the NMRC Office of Research Administration and Human Subjects Protections Program. All potential study

subjects provided written, informed consent before screening and enrollment and had to pass an assessment of understanding.

2) CHMI Study: Twelve subjects were recruited for the CHMI study along with two alternates to participate in the infection control arm of a recent U.S. Navy malaria vaccine clinical trial [18]. All subjects were adults between the ages of 18 and 50 years old who satisfied stringent screening requirements using medical history, physical examination, electrocardiogram, and laboratory testing to be considered "healthy" and fit for malaria challenge. Any subject with prior malaria exposure was excluded. Cardiac risk screening was conducted to identify and exclude individuals at moderate or high risk of developing symptomatic coronary artery disease. All subjects were given wearable devices and instructions for use upon enrollment. The study protocol and use of research data were approved by the NMRC IRB in compliance with all applicable federal regulations governing the protection of human subjects.

3) Control Study: Nine healthy adult subjects conforming to the same inclusion criteria yet not involved with the challenge study served as controls. These subjects were asked to carry out the same device usage instructions as those in the CHMI study, and followed the same study procedures. The wearable data from these subjects were used for method development and to assess the overall specificity and the false-positive rate per week of the *2B-Healthy* algorithm.

B. Wearable Devices and Software

Subjects were given a Samsung Gear S3 smartwatch paired to a Samsung Galaxy S7 smartphone (Samsung Electronics, Inc., Ridgefield Park, NJ). The smartwatch collected heart rate, skin temperature, and body acceleration data. The Samsung Gear S3 was selected as the wearable device both because of its sensor accuracy (see below) and because it provided an Application Programming Interface that allowed for direct access to the sensor data, without requiring third-party involvement or upload to a cloud-based service. Due to the limited data storage capacity on the smartwatch, data collected by the watch were automatically and wirelessly transferred via Bluetooth to the paired smartphone. Subjects were instructed to wear the smartwatch for at least 12 hours a day with a consistent daily usage pattern, from the start of the study, approximately 2 weeks pre-challenge, to the study conclusion, 4 weeks post-challenge. As configured for the study, the smartwatch battery life was less than the 30 hours in the default configuration, and subjects were instructed to charge their smartwatch once a day. Subjects were instructed to keep the smartphone at home so that it would sync once daily with the smartwatch to download the data. Finally, subjects were instructed to come to the study site once a week to have the wearable data stored in the smartphone downloaded to a laptop to assess user compliance, provide feedback on any device-related issues, and ensure that data were saved in the event of loss or damage of a device.

The Samsung Gear S3 smartwatch and Samsung Galaxy S7 smartphone are COTS devices, but they were specially configured for this study. An in-house-developed app was loaded to both the smartwatch and the smartphone. This app set the photoplethysmography (PPG) sensor sampling rate to the maximum 100 Hz to ensure the highest accuracy for heart-rate data collection. This app also automatically and wirelessly transferred data from the smartwatch to the smartphone via Bluetooth when the devices were in range. Finally, the app on the smartphone included a user interface that allowed subjects to see if their smartwatch was functioning appropriately and provide visual feedback on the vital signs being recorded.

C. Assessment of Wearable-Sensor Accuracy

We carried out an in-house assessment of the accuracy of the PPG, temperature, and 3-axis accelerometer sensors on the Samsung Gear S3 by comparing their values against clinical-grade (gold-standard) instruments, specifically, the Polar H7 heart-rate sensor (Polar Electronics Co. Ltd., Guangzhou, China), the iButton temperature sensor (iButtonLink LLC, Whitewater, WI), and the ActiGraph wGT3X-BT actigraphy device (Acti-Watch LLC, Pensacola, FL). For each sensor, we had subjects wear both the Samsung Gear S3 and the gold-standard device simultaneously under every day, ambulatory conditions. We then analyzed the sensor data from both the Samsung Gear S3 and the gold-standard devices to measure the accuracy of the wearable sensors. The results of this assessment are provided in Supplemental Table S1. Briefly, we found that the median error magnitude (i.e., the median of the absolute error) between the Samsung Gear S3 sensor and the gold standard was <5 bpm for heart rate, <5 °C for skin temperature, and $<1\%$ for the actigraphy data. In terms of activity levels, the median errors between the Samsung Gear S3 and the gold-standard heart rate measurements were 1 bpm, 4 bpm, and 9 bpm for resting, low, and moderate activity levels, respectively. In addition to these benchmark tests, further in-house testing revealed that wrist-based skin-temperature data from the smartwatch were highly sensitive to transient factors, such as environment, clothing, and posture, and were thus considered too unreliable for use in this study.

D. 2B-Healthy Algorithm

We developed the *2B-Healthy* algorithm to analyze heart-rate and actigraphy data and provide a real-time estimate of the probability of infection. This algorithm consisted of two steps: *1)* detection of aberrant heart-rate (AHR) patterns and *2)* real-time estimation of probability of infection through the accumulation of elevated AHR frequencies over time.

1) Data Pre-Processing: The raw heart-rate data were collected at a sampling rate of 1 Hz, and then averaged in non-overlapping 5-second windows. A moving average was then taken using a 15-minute window. The activity data were collected by a 3-axis accelerometer at 25 Hz, and averaged in a non-overlapping 5-second window along each axis. We then converted the 5-second mean acceleration along the three axes into a magnitude of acceleration vector. This value was then used to calculate a metabolic equivalent level (MET level) using an in-house algorithm [3] based on a previously published algorithm [19]. The in-house algorithm showed high correlation with the original version, but was developed based on 125 hours

of activity data collected in-house that was carefully annotated by activity (e.g., "sleeping," "office work," "walking," and "exercising"). A moving average of the MET levels was then taken with a 15-minute window, resulting in pre-processed heart rates at 5-second intervals paired with MET levels at the same time intervals.

2) AHR Detection: In order to identify AHRs, the algorithm first generated an individualized baseline distribution of "healthy" (non-infection) heart rates at different times of day and levels of activity. The baseline data were obtained from the pre-challenge time period for CHMI subjects, and from the first two weeks of data collection for the control subjects. The levels of activity were obtained by converting the raw acceleration data into MET values [3] and then mapping these values into three activity levels: "resting," "low," and "moderate." To define the MET cutoff values associated with each activity level, we used an activity-annotated dataset, where the acceleration data and associated activity annotation were previously collected using the Samsung Gear S3 smartwatch. Based on these activity-annotated data, we defined MET values of 0.50 to 0.75, which included annotated activities such as "sleep," as an activity level of resting; MET values of 0.75 to 1.00, which included annotated activities such as "office work," as an activity level of low; and MET values of 1.00 to 1.25, which included annotated activities such as "walking," as an activity level of moderate. Supplemental Table S2 provides additional information. Higher activity levels were excluded because too little data were collected for reliable baseline estimates.

For each subject, baseline distributions at different times of day (t_k) about a window of time $(\pm w)$, at each of three activity levels (a_j) were determined. For a given query period, the average heart rate over a 30-minute window $[HR(t_k, a_j)]$ was calculated from the pre-processed heart-rate data annotated by the corresponding time of day and activity level during that time window. This average heart rate was then converted to a *HR Score* for a time window $t_k \pm w$ and activity level a_j, based on the mean μ and standard deviation σ of the baseline distributions as described in (1):

$$HR\ Score = \frac{HR(t_k, a_j) - \mu(t_k \pm w, a_j)}{\sigma(t_k \pm w, a_j)} \qquad (1)$$

We defined a minimum and maximum cutoff for *HR Score*, outside of which a heart rate was considered to be an AHR. These cutoffs were based on the upper and lower bounds corresponding to 95th percentile limits of the baseline distribution and were calculated by fitting a Weibull distribution to the baseline distributions for each subject, at each activity level. Parameterization, in terms of finding appropriate values for the window of time (w), heart-rate average window (30 minutes), and threshold cutoffs (95th percentile), was determined manually with the goal of obtaining *HR Scores* and AHR classifications that were robust to week-to-week variations in baseline data, while being sensitive to clusters of abnormally elevated or decreased heart rates. These analyses were carried out using both simulated and real-world wearable-collected data. One important advantage to using an activity-level specific, percentile-based approach

to detecting AHRs is that it accounts for the greater noise in heart rate data at higher activity levels by comparing all heart rate data to baseline data collected at that same activity level. Thus, as long as the noise with respect to activity level is comparable between the baseline and detection periods, increased noise at higher activity levels will not result in higher AHR frequencies.

3) Real-Time Estimate of Infection Probability: *2B-Healthy* relies on a Bayesian-based model for estimating real-time infection probability by accumulating evidence of AHRs over time. The real-time probability of infection at time t (P_t) about a time window of 30 minutes is a function of the number of aberrant (n) and normal (m) heart rates observed in the time window around time t, along with the prior probability of infection at the previous time window (P_{t-1}), as described in (2). Additional fixed parameters include the probability of observing AHRs in healthy (P_h) and infected (P_i) individuals. There is also a "forgetting" factor (α), with $\alpha < 1.0$, which causes the contributions of prior probabilities to the current probability to decay with time, creating a recency bias to P_t. The equation is based on Bayes theorem and estimates the posterior probability of infection P_t, at time t, based on a prior probability of infection, the likelihood of observing normal and aberrant heart rates in healthy and infected individuals, and the new evidence–the number of newly observed normal and aberrant heart rates at time t. In the extreme case where no heart rates are observed, the posterior probability is equivalent to α multiplied by the prior probability, leading to a slight decrease in the prior probability of infection reflecting the lack of recent data. In cases where no normal or aberrant heart rates are observed, P_t will increase or decrease, respectively, based on the parameters P_h and P_i and the prior probability.

$$P_t = \frac{1}{1 + \left(\frac{P_h}{P_i}\right)^n \left(\frac{1 - P_h}{1 - P_i}\right)^m \left(\frac{1}{\alpha P_{t-1}} - 1\right)} \qquad (2)$$

Values for fixed parameters P_h and P_i were set based on the goal of obtaining the maximum separation in peak infection probability between wearable data obtained from CHMI subjects (during the incubation phase) and control subjects. We did not use a supervised machine-learning approach to parameter fitting due to the small sample size of the study and the high risk of over-fitting. Instead, we manually set the parameters based on key desired behaviors of the model: *1)* $P_i > P_h$, based on the assumption that infection leads to an increase in AHR frequency; *2)* $P_h = 0.05$, reflecting the AHR thresholds set about the limits of the 95th percentile of the baseline (healthy) distribution; and *3)* AHRs observed more than 5 days prior should be ignored based on the assumption that early stages of infection rarely exceed this time span. We adjusted the α parameter so that the contribution of AHRs observed 5 days or earlier becomes zero. Since P_i is not known, we adjusted its value, as a ratio of P_h, and observed its effect on P_t; this effectively adjusted the sensitivity of the prediction model to the presence of AHRs and is the single-most important parameter in the model.
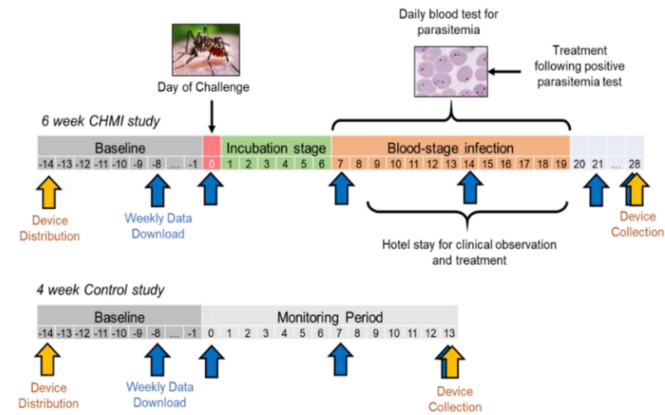
Fig. 1. Overview of wearable CHMI study. Subjects were enrolled 2 weeks pre-challenge and given wearable devices. Subjects were challenged on Day 0 and checked into a hotel on Day 9 for monitoring by clinical staff. A daily log of malaria-related symptoms was recorded, and daily blood tests for parasitemia were carried out starting from Day 7 until parasitemia was detected or the study was concluded. Antimalarial treatment was administered immediately following a positive parasitemia test or at Day 28, prophylactically. Subjects were instructed to wear the wearable device for at least 12 hours/day and to come in once weekly to upload their wearable data to a study laptop.

E. Measures for User Compliance and Algorithm Performance

We measured user compliance in terms of the percentage of days over the course of the study during which the user wore the device for at least 12 hours. Device reliability was measured in terms of the percentage of devices that had to be replaced at least once. *2B-Healthy* infection-detection performance was measured in three ways: *1)* the maximum probability of infection estimated during the incubation phase for a CHMI subject (P_{max}); *2)* the time to detection prior to parasitemia (T_d), as detected by blood smear; and *3)* time to detection prior to onset of symptoms (T_s). We assessed the sensitivity of *2B-Healthy* based on the percentage of CHMI subjects where infection was detected early (during incubation) and at any time following challenge, as determined by an infection probability of $\geq$50%. We calculated the specificity of *2B-Healthy* based on the percentage of control subjects who had an estimated infection probability of $\geq$50%, and calculated the false-positive rate as the frequency that an individual in the control group would have an infection probability $\geq$50% in a 1-week (non-overlapping) span.

III. RESULTS

A. CHMI Challenge

Twelve subjects received the wearable devices upon enrollment, approximately 2 weeks pre-challenge. All of the subjects were part of an infection control arm of a malaria-vaccine study [18] in which the subjects were challenged but not vaccinated. The overall CHMI study schedule is shown in Fig. 1. Briefly, subjects were challenged on Day 0 via bites from *Plasmodium falciparum*-infected mosquitoes as described by Sklar *et al.* [18]. From post-challenge Day 7, subjects were regularly
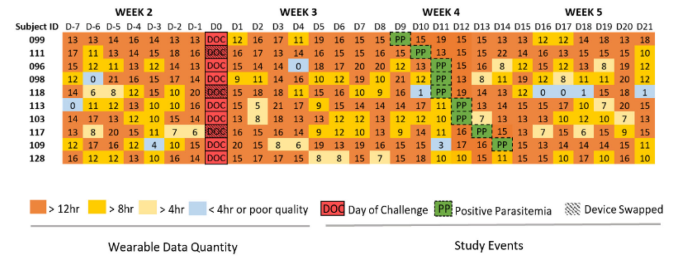

Fig. 2. Hours of wearable data collected in the CHMI study from Weeks 2–5. The daily quantity of wearable data collected from each CHMI subject is shown from 1 week pre-challenge to 3 weeks post-challenge. Each cell shows the number of hours of data collected that day and is colored corresponding to >12 hours (orange), >8 hours (yellow), >4 hours (light yellow), or <4 hours (light blue) of "good-quality" data, defined as within physiological range for 90% of the time span of wearable data collected. The day of challenge (DOC, red) and the first day of positive parasitemia (PP, green) are shown along with device reliability (device replaced). Subjects are sorted by time to parasitemia. One subject (Subject 128) was challenged but did not show blood-stage parasitemia.

monitored for malaria-related symptoms and subjected to daily blood-smear tests for blood-stage parasitemia until either parasitemia was detected or the study was concluded. On Day 9, subjects were checked-in to a hotel so that they could be monitored daily by the clinical team; during this period, they were free to carry out their regular routine, such as going to work, during the day. Following a positive parasitemia test, an FDA-approved antimalarial drug (Malarone^{TM} or Coartem^{TM}) was immediately administered. Throughout the study, subjects were asked to come to the clinical site (or hotel) once a week to download the wearable data from the smartphone to a study laptop.

Overall, all 12 subjects were challenged on Day 0, and 11 of 12 CHMI subjects subsequently developed blood-stage parasitemia, between 9 and 14 days post-challenge (average 11.7 days), as determined by blood smear. Four of 12 subjects reported at least one moderate or severe malaria-related symptom within 2 weeks of challenge, with an average time to onset of the first symptom at 9.5 days post-challenge. All subjects were given antimalarial treatment immediately upon receiving their first positive parasitemia test, except the one subject who did not present blood-stage parasitemia and was treated presumptively on Day 28. All subjects showed three consecutive daily negative blood-stage parasitemia tests prior to the conclusion of the study, indicating that the parasite had been successfully cleared by treatment.

B. Device Usage and Compliance

Subjects were assessed for device usage compliance as well as device reliability. Two of the 12 CHMI subjects showed compliance below 50%, were deemed to have collected insufficient wearable data for further analysis, and were excluded from the remainder of the study. We focused our compliance assessment over the time span of Week 2 through Week 5, to include the week pre-challenge to 3 weeks post-challenge (Fig. 2), where any infection-related effects would be observable. Of the 10 subjects, overall compliance for device usage of at least 12

hours/day during this span was 72%, and overall compliance for at least 8 hours/day was 92%. The major factor limiting higher compliance and wear time was likely the fact that the Samsung Gear S3 smartwatch, as configured in the study, needed to be charged daily for 2-3 hours, limiting the time that subjects could be wearing the devices on a given day. Three of the 10 subjects had to have their wearable device replaced due to reliability issues, for an overall reliability of 67%. For comparison, data collection for control subjects is shown in Supplemental Fig. S1. The main issue with reliability was an occasional but previously documented malfunction of the Samsung Gear S3 smartwatch that was found in the firmware version used in the study (Tizen version 3.0.0.2 and build number R760XXU2CRH1), which altered the frequency rate and values of some parameters and quickly drained the battery. In these cases, the device was replaced by another study device.

C. AHR Detection

We combined the raw heart-rate data and the actigraphy data to generate a *HR Score* that took into account circadian rhythm and activity, and then implemented an AHR classification scheme whereby any *HR Score* outside the thresholds defined by the 95th percentile limits of the corresponding baseline distributions was classified as "aberrant" and all others were classified as "normal." Fig. 3 shows representative heart-rate (Fig. 3(a)) and actigraphy data (Fig. 3(b)) collected from a single subject over a 7-day period that were converted into average heart rates and finally into *HR Scores* across all times of day, for each activity level, with those of low activity level shown in Fig. 3(c) and Fig. 3(d), respectively.

The overall frequency of AHR (relative to total number of cheart-rate data points) for both the baseline and the post-baseline monitoring period is shown in Table I for CHMI and control subjects. Because the baseline data were used to define the 95th percentile *HR Score* thresholds that define AHRs, it is expected that the AHR frequency during the baseline period should be approximately 5%.

Overall, we found that the average AHR frequency in the baseline period was 5.6% for CHMI subjects and 4.9% for control subjects. For CHMI subjects during the post-challenge monitoring period, the average AHR frequency rose ∼2-fold to 11.6%, while in the control subjects, the average AHR frequency during the post-baseline monitoring period was 8.0%. Both CHMI and control subjects showed an increase in AHR frequency from baseline ($P < 0.01$ and $P < 0.05$, respectively). However, CHMI subjects showed a significantly larger increase in AHR frequency than control subjects ($P < 0.05$), suggesting that increased AHR frequency may be indicative of infection. Furthermore, we found that AHRs were observed at all three activity levels, in both CHMI and control subjects, roughly corresponding to the amount of wearable data collected at each level, with the most AHRs detected at the "low" activity level (Supplemental Fig. S2). For some subjects, we observed no AHRs at the "resting" activity level because there were insufficient wearable data collected at this activity level for those subjects.
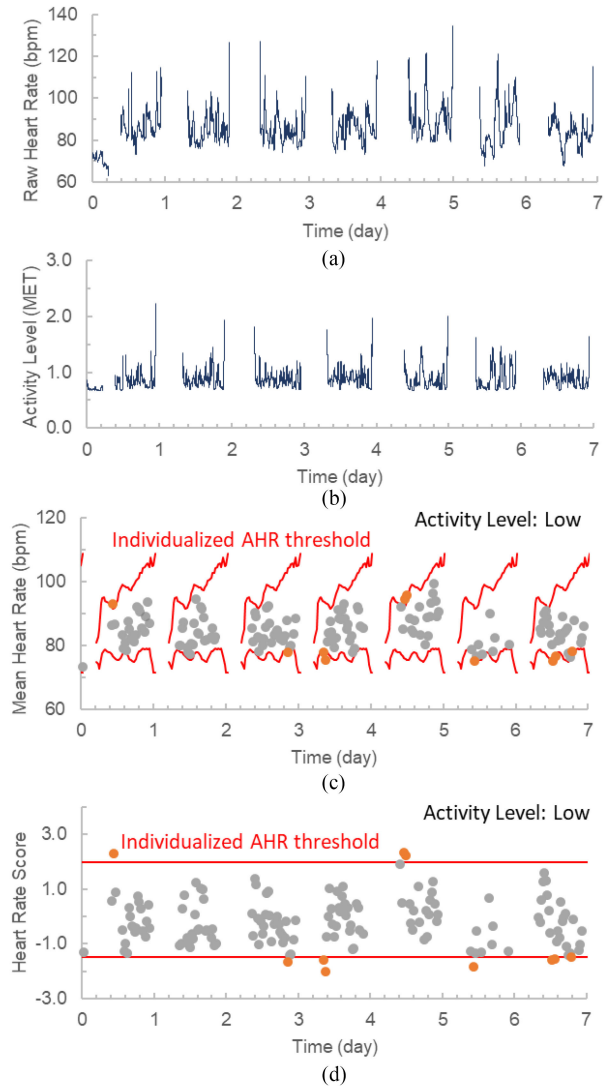

Fig. 3. Aberrant heart rate detection in *2B-Healthy* for a representative 7-day window. (a) Raw heart-rate data, averaged in 30-second intervals. (b) Raw activity data represented by the magnitude of acceleration vector averaged in 30-second intervals. (c) Activity data were used to stratify the heart-rate data by three activity levels (resting, low, and moderate), and the average heart rate in 30-minute non-overlapping windows was calculated for each activity level (low activity level shown). An individualized AHR threshold (red line), determined from baseline data, defines normal (gray) and aberrant heart rates (AHR; orange). (d) The heart-rate data were normalized across all times of day into a *HR Score*, for each activity level (low activity level shown), with the AHR threshold (red line), normal heart rates (gray), and AHRs (orange) shown.

D. Estimating Probability of Infection

We developed a Bayesian-based model for estimating the probability of infection at a given time point (P_t) from AHR frequency data. The model contains three main parameters: *1)* P_h, the probability of observing an AHR in a healthy (non-infected) individual; *2)* P_i, the probability of observing an AHR in an infected individual; and *3)* α, a decay factor that biases P_t to more recent observations.

Subject	Baseline			Monitoring Period		
	#AHRs	#NHRs	%AHR	#AHRs	#NHRs	%AHR
CHMI						
96	19	363	5.0	48	537	8.2
98	13	166	7.3	75	399	15.8
99	14	245	5.4	60	497	10.8
103	5	141	3.4	60	325	15.6
109	23	410	5.3	70	900	7.2
111	25	488	4.9	75	702	9.7
113	22	306	6.7	138	657	17.4
117	23	358	6.0	112	640	14.9
118	27	402	6.3	57	501	10.2
128	12	190	5.9	52	762	6.4
Average	**18**	**307**	**5.6**	**75**	**592**	**11.6**
Control						
1	20	407	4.7	24	457	5.0
2	8	271	2.9	33	244	11.9
4	15	385	3.8	23	256	8.2
5	11	189	5.5	10	230	4.2
6	31	433	6.7	55	443	11.0
7	15	307	4.7	14	244	5.4
8	26	411	5.9	54	493	9.9
9	20	367	5.2	27	287	8.6
Average	**18**	**346**	**4.9**	**30**	**332**	**8.0**

AHRs = abnormal heart rates; NHRs = normal heart rates

Fig. 4 shows the effect of adjusting P_i, as a ratio of P_h, on the maximum observed P_t during the incubation period (CHMI subjects) and total monitoring period (control subjects). We found that there was separation in the maximum observed P_t between CHMI and control subjects across a range of values for P_i/P_h from 1.6 to 1.9. For most CHMI subjects (6 of 9 that achieved parasitemia), their maximum P_t exceeded 40% in this range of P_i/P_h. Based on the observed separation, we selected a P_i of $1.7*P_h$ in order to calculate P_t, and a P_t threshold of 50% above which a subject would be classified as "infected" in the *2B-Healthy* algorithm. With P_i/P_h at 1.7, we adjusted the "forgetting" factor α to be 0.988 based on simulated heart-rate data.

E. 2B-Healthy Infection-Detection Performance

Table II summarizes the infection-detection performance in CHMI and control subjects. Infection was detected in seven of nine CHMI subjects that developed parasitemia; in six of nine subjects it was detected prior to the detection of parasitemia by blood smear, with an average T_d of 6.4 days. Four CHMI subjects reported moderate or severe CHMI-related symptoms within 2 weeks post-challenge; in all four cases infection was detected ($P_t > 50\%$), with an average T_s of 3.3 days. In control subjects, P_t exceeded the 50% threshold in two of eight cases, over a
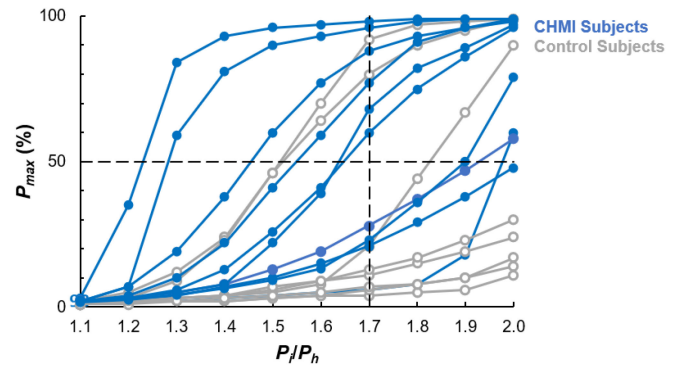

Fig. 4. Probability of infection as a function of P_i. We adjusted the parameter for the expected probability of observing an AHR in an infected individual (P_i) as a ratio of the probability of an AHR in a healthy individual ($P_h = 0.05$) and assessed its effect on the maximum estimated probability of infection (P_{max}) during the incubation period (CHMI subjects, blue) or total monitoring period (control subjects, gray). Based on the separation between CHMI and control subjects, we selected a P_i value of $1.7*P_h$ to calculate P_t and an infection probability threshold of 50% for determining that a subject is "infected."

CHMI Subject	P_{max} (%)	T_d (days)	T_s (days) [Severity]	Control Subject	P_{max} (%)
117	99	10.0		004	4
113	96	4.6	5.6 [Moderate]	007	5
111	88	4.4		001	11
103	77	8.1		005	13
98	68	3.5	4.5 [Severe]	009	18
118	83	7.5	5.5 [Moderate]	002	35
96	29	-1.3	-2.3 [Severe]	008	76
109	23			006	96
99	6				
128*	17				

*CHMI subject 128 was challenged but did not achieve parasitemia.

combined monitoring of 32 weeks. Overall, the sensitivity for infection detection was 78% (7 of 9), the sensitivity for infection detection prior to parasitemia was 67% (6 of 9), and the false-positive rate based on control subjects was 6% per week, for an overall specificity of 75%.

Subject 113 (Fig. 5(a)) serves as a representative example for strong performance by the *2B-Healthy* algorithm, as measured by P_{max}. This subject experienced moderate symptoms (headache) on Day 11 and reached blood-stage parasitemia by Day 12. P_t for this subject exceeded the 50% threshold at Day 7, 4.6 days prior to blood-stage parasitemia, reaching a P_{max} of 96%. Subject 98 (Fig. 5(b)) represented the median performance of the *2B-Healthy* algorithm as measured by P_{max}. This subject had blood-stage parasitemia by Day 11 and experienced severe symptoms (fatigue) and moderate symptoms (malaise and myalgia) on Day 12. P_t for this subject reached the 50% threshold at Day 7, 3.5 days prior to parasitemia. Finally, subject 96 (Fig. 5(c)) represents a poor prediction by the *2B-Healthy*
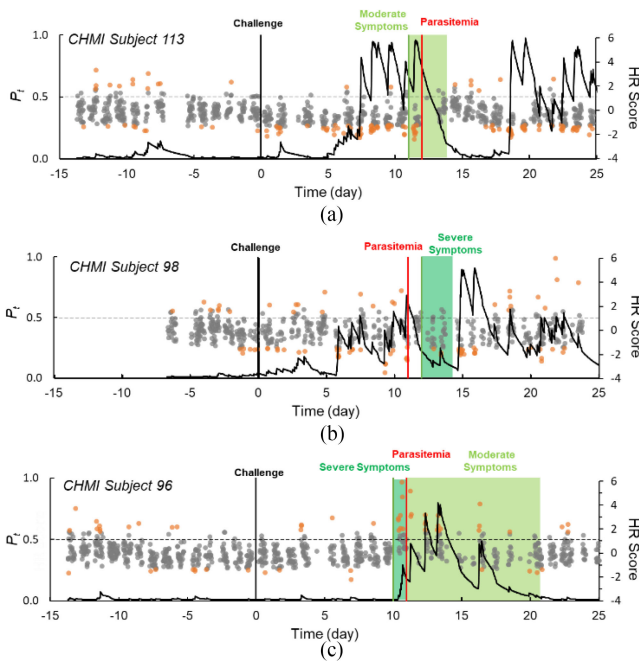
Fig. 5. *2B-Healthy* infection detection for three representative CHMI subjects. P_t (black line) and heart-rate *(HR) Score* (aberrant, orange; normal, gray) are shown over time from the pre-infection (Day -15 to Day 0) to post-challenge (Day 1 to Day 25) periods. Time points for malaria challenge (black vertical line), parasitemia as detected by blood smear (red vertical line), and onset and duration of moderate (light green) and severe (dark green) symptoms are shown for three CHMI subjects: subject 113 (strong detection performance, (a)), subject 98 (average detection performance, (b)), and subject 96 (weak detection performance, (c)).
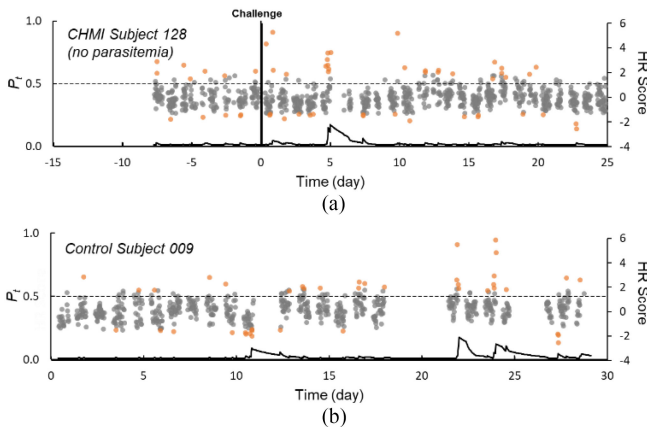

Fig. 6. *2B-Healthy* infection detection in subjects without infection or parasitemia. P_t (black line) and heart-rate *(HR) Score* (aberrant, orange; normal, gray) are shown over time. CHMI subject 128 (a) was challenged on Day 0 (black vertical line) but did not develop parasitemia. Control subject 009 (b) represents the median *2B-Healthy* performance in non-infected subjects.

algorithm. This subject presented blood-stage parasitemia on Day 11 and severe symptom (fever) onset on Day 10. *2B-Healthy* did not show P_t greater than 50% during the incubation period, but did reach that threshold on Day 12, 1.3 days after blood-stage parasitemia was detected by blood smear.

Fig. 6 shows two examples of *2B-Healthy* infection-detection performance in subjects that did not have infection. CHMI subject 128 underwent challenge but did not develop parasitemia. Consistent with the blood-based diagnostic result and the lack of malaria-related symptoms, *2B-Healthy* also did not report a P_t value exceeding the 50% threshold at any point during the post-challenge monitoring period. Likewise, in control subject 009, the median subject in the control group in terms of prediction performance as assessed by P_{max}, the P_t reported by *2B-Healthy* was close to zero for almost the entire duration of the monitoring period.

IV. DISCUSSION

To our knowledge, this study represents the first wearable-based physiological monitoring study carried out using a controlled human-infection model. The CHMI model is ideal for studying early infection detection because it is *1)* a well-established disease model; *2)* presents symptoms common to many infectious diseases, such as fever, headache, and malaise; *3)* has a lengthy incubation period; and *4)* produces a range of outcomes both in time course and severity. It is important to note that *Plasmodium falciparum* malaria, undiagnosed and untreated, is life-threatening and can persist for weeks after the onset of parasitemia. The blood-smear diagnostic is able to detect infection very early in the disease progression, often before the onset of symptoms and sufficiently early to ensure complete recovery following treatment. This makes the blood-smear diagnostic a clinical gold standard for early infection detection for comparison of wearable-based methods.

In this study, we used a COTS Samsung Gear S3 smartwatch to collect heart rate, skin temperature, and 3-axis accelerometer data from 12 subjects as they underwent CHMI. During the study, subjects were able to live at home and go about their daily routine, except for a ∼1-week-long period when they were required to spend their nights in a hotel room. We hypothesized that heart-rate data, although highly variable, could provide a reliable measure of physiological state once factors such as circadian rhythm and physical activity were accounted for. Thus, we developed *2B-Healthy*, a method that generates an individualized baseline heart-rate distribution, accounting for circadian rhythm and activity for each individual, which it then uses to identify AHR patterns during a monitoring period and estimate a real-time probability of infection.

Our findings showed that most CHMI subjects (6 of 9) who reached parasitemia showed a significant increase (∼2-fold) in AHR frequency following challenge, compared to control subjects (2 of 8) and the one CHMI subject who did not reach parasitemia. We expected that infection might lead to an *elevated* heart rate, relative to baseline, based on the long-observed relationship of an elevated heart rate associated with fever [20]. Interestingly, however, we did not observe a consistent pattern with respect to the direction of the AHRs (high vs. low) in CHMI subjects, suggesting that this infection-induced increase in AHR frequency may be the result of behavioral factors, such as changes in the intensity, manner, or types of activities carried out during the day not captured by activity levels, as well as

physiological factors that directly affect heart rate. One example of behavioral factors that could lead to decreased heart rates could be that as an individual begins to feel ill, they opt to engage in less active or more restful activities *within* a particular activity level than they normally would, such as napping instead of reading while in bed, or sitting instead of standing while doing chores. Interestingly, a recent COVID-19 wearable study by Radin *et al.* [21] observed both tachycardia and bradycardia following infection, suggesting that the expected relationship between infection and tachycardia may not be so straightforward in wearable data where individuals may opt to engage in a wide range of activities.

Overall, using a $P_t >50\%$ threshold for detecting infection, we found that *2B-Healthy* was able to detect infection in seven of nine CHMI subjects that reached parasitemia, for an overall sensitivity of 78%. More strikingly, it was able to detect infection prior to the blood-smear diagnosis in six of nine cases, for a sensitivity of 67% with respect to this clinical gold standard for early malaria detection. In some cases, P_t increased again post-parasitemia during the treatment phase, which may reflect persistent malaria symptoms following parasitemia or possibly side effects of the 3-day course of the anti-malarial treatment (Malarone$^{\mathrm{TM}}$ or Coartem$^{\mathrm{TM}}$). Among eight control subjects, *2B-Healthy* detected infection in two cases, for a false-positive rate of 6% per week (2 weeks out of a total 33 non-overlapping weeks of data) and an overall specificity of 75% (6 out of 8 control subjects showed no infection). While we refer to infection detection in these control subjects as false positives, they cannot, with certainty, be described as false detections, because it is not known whether another infection (for example, the cold or flu) or other adverse health event led to an increase in the AHR frequency.

The *2B-Healthy* algorithm is based on detection of heart rate abnormalities, a non-specific physiological characteristic that may pose challenges to real-world implementation. Because a wide range of infections can lead to heart rate abnormalities, the algorithm's intended use is for *screening* rather than diagnostics. However, numerous non-infection-related factors can also cause heart rate abnormalities which could undermine the algorithm's utility as a screening tool. In this study, using the control group, we observed a rate of possible non-infection-related heart rate abnormalities of 6% per week. While this false-positive rate may be too high for the practical use of *2B-Healthy* as a general screening tool, it is sufficiently low for effective screening in situations where infectious disease prevalence is anticipated to be moderate to high. For example, with a 6% weekly false-positive rate and 70% sensitivity, and a screening strategy where a positive *2B-Healthy* detection results in follow-up clinical diagnostic, if the prevalence of disease is 1 to 2%, *2B-Healthy* would be able to identify 70% of infected individuals using only ~7% of the testing resources required for a universal weekly diagnostic screening. Additional study is needed to further assess the false-positive rate of *2B-Healthy* in real-world settings as well as to explore the use of additional wearable-based physiological measures, such as temperature and blood oxygen saturation, which could be combined with heart rate to increase specificity sufficiently for general use.

We also found several additional limitations and challenges in the present study. The sample size for this study is sufficient for assessing feasibility, but not for rigorously evaluating infection-detection sensitivity or specificity. A second limitation was the relatively limited baseline data collected for each individual in the study. *Post hoc* analysis showed that reliable baseline estimates for heart rates require 2-3 weeks of data (data not shown); most subjects in this study provided 1-2 weeks worth of baseline data. This is in contrast to real-world use of wearable devices, where daily use is common and a large quantity of baseline data can be readily obtained. A third limitation was the battery life of the Samsung Gear S3 smartwatch as it was configured for this study, necessitating 2-3 hours of daily charging. Because of this limitation, we opted for day-time wear of the device to better reflect typical ambulatory use of wearable devices, which resulted in a lack of night-time data and a reduced overall daily quantity of wearable data to approximately 12 hours/day. Another challenge in this study was the stability and reliability of the COTS wearable device; these are typically consumer electronics products with short life cycles and frequent firmware updates that can affect key device characteristics, such as sensor sampling rate, signal characteristics, and battery life. Further testing and validation of *2B-Healthy* infection-detection performance is needed in independent follow-up studies.

There have been a few observational studies using COTS smartwatch-based wearable devices to detect infection during the COVID-19 pandemic, most finding that infection is associated with abnormal resting heart rate [6], [7], [9], respiration rate [9], and sleep and activity [6], [7]. A separate study using a ring-based wearable device reported abnormal heart rate and skin temperature readings in infected subjects [8]. Because time of exposure in these studies is not known, early detection is difficult to evaluate. Nonetheless, two studies report observing physiological changes prior to the onset of symptoms in some percentage of infected subjects [6], [8]. Comparison of these studies with the present one is difficult because of differences in methodology and disease, but the sensitivity and specificity observed in these studies are comparable to what we report here: 50-70% sensitivity at 80% specificity [7], [9], with pre-symptomatic or early detection in 60% of cases in retrospective analysis [6]. Although different infectious diseases have different time courses and severities, the physiological data collected in all of these studies–heart rate, skin temperature, and activity–are highly generalizable and likely non-specific to any particular disease. As such, further studies using wearable devices across a range of infectious diseases in observational studies as well as controlled infection models may provide broad insights for advancing wearable-based infection-detection technology.

V. CONCLUSION

We conducted a wearable device feasibility study using the COTS Samsung Gear S3 smartwatch to monitor heart rate, skin temperature, and body acceleration data in subjects as they underwent CHMI. We developed *2B-Healthy*, an algorithm that analyzes heart rate and activity data, identifies aberrant

heart-rate patterns, and estimates in real time a probability of infection. Our findings showed that *2B-Healthy* was able to detect infection prior to the clinical-grade blood-based diagnostic with a sensitivity of 67%, with a false-positive rate of 6% per week. While this false-positive rate may be too high for general population use where it might greatly exceed the prevalence rate of infection, in circumstances with the potential for high prevalence or susceptibility this could serve as a vital screening tool for detecting and treating infections. Such examples of high disease prevalence ($>10\%$) include diarrhea or dysentery among international travelers, malaria during the rainy season in a country where it is endemic, influenza and adenovirus infections in basic training for military personnel, and even health-care settings during an outbreak or pandemic response.

The present study is unique among wearable infection studies in several key aspects: it uses a controlled infection model in order to precisely establish the time of exposure and onset of infection; it uses a Bayesian-based algorithm that accounts for circadian rhythm and activity level to estimate an individualized probability of infection and enable real-time inferences; and the *2B-Healthy* algorithm is device-agnostic and disease non-specific, allowing its integration with any wearable device that accurately measures heart rate and activity for any infection event that generates abnormalities in heart rate patterns. These findings underscore the value of using controlled infection studies to develop early infection-detection technologies, demonstrate the feasibility of using wearable devices to detect infection, and highlight the potential use of this technology in screening efforts. Further independent validation of *2B-Healthy* in follow-up, prospective clinical and observational studies will be critical to evaluate its performance as a wearable-based early infection-detection platform.

DISCLAIMER

ETHICS STATEMENT

The study protocol for this clinical trial and use of the research data were approved by the Walter Reed Army Institute of Research and the Navy Medical Research Center Institutional Review Boards, in compliance with all applicable Federal Regulations governing protection of human subjects. All study subjects gave written, informed consent. The study protocol for the control study was reviewed and approved by the Human Research Protection Office, Ft. Detrick, Maryland.

AUTHOR CONTRIBUTIONS

SC, JEE, and JR designed the study. CD and JMS carried out the CHMI clinical study and provided clinical data, such as malaria diagnostics and symptomology. SC carried out the wearable data collection for the CHMI subjects. SH carried out the wearable data collection for the control subjects. KK developed the watch and phone applications for data collection. SC, CY, RL, SH, and JR carried out the data analysis and development of the *2B-Healthy* algorithm. SC and JR wrote the manuscript. All authors have reviewed the manuscript and approved the submitted version.

CONFLICT OF INTEREST

The authors declare the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

DATA AVAILABILITY STATEMENT

The data from the studies will be made available following a written request to the corresponding author, along with a summary of the planned research.

ACKNOWLEDGMENT

REFERENCES

[1] X. Li *et al.*, "Digital health: Tracking physiomes and activity using wearable biosensors reveals useful health-related information," *PLOS Biol.*, vol. 15, no. 1, Jan. 2017, Art. no. e2001402.

[2] S. Laxminarayan *et al.*, "Human core temperature prediction for heat-injury prevention," *IEEE J. Biomed. Health. Inform.*, vol. 19, no. 3, pp. 883–891, May 2015.

[3] S. Laxminarayan *et al.*, "Individualized estimation of human core body temperature using noninvasive measurements," *J. Appl. Physiol.*, vol. 124, no. 6, pp. 1387–1402, Jun. 2018.

[4] A. Gani *et al.*, "Universal glucose models for predicting subcutaneous glucose concentration in humans," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 157–165, Jan. 2010.

[5] J. Reifman *et al.*, "*2B-Alert* App: A mobile application for real-time individualized prediction of alertness," *J. Sleep Res.*, vol. 28, no. 2, pp. e12725, Apr. 2019.

[6] T. Mishra *et al.*, "Pre-symptomatic detection of COVID-19 from smartwatch data," *Nature Biomed. Eng.*, vol. 4, no. 12, pp. 1208–1220, Dec. 2020.

[7] G. Quer *et al.*, "Wearable sensor data and self-reported symptoms for COVID-19 detection," *Nature Med.*, vol. 27, no. 1, pp. 73–77, Jan. 2021.

[8] B. L. Smarr *et al.*, "Feasibility of continuous fever monitoring using wearable devices," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 21640.

[9] A. Natarajan, H. W. Su, and C. Heneghan, "Assessment of physiological signs associated with COVID-19 measured using wearable devices," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 156, Nov. 2020.

[10] J. D. Chulay *et al.*, "Malaria transmitted to humans by mosquitoes infected from cultured *Plasmodium falciparum*," *Am. J. Trop. Med. Hyg.*, vol. 35, no. 1, pp. 66–68, Jan. 1986.

[11] World Health Organization, *World Malaria Report*, Geneva, Switzerland: WHO, 2018.

[12] M. Spring, M. Polhemus, and C. Ockenhouse, "Controlled human malaria infection," *J. Infect. Dis.*, vol. 209, no. Suppl 2, pp. S40–S45, 2014.

[13] M. C. Langenberg *et al.*, "Controlled human malaria infection with graded numbers of *Plasmodium falciparum* NF135.C10- or NF166.C8-infected mosquitoes," *Am. J. Trop. Med. Hyg.*, vol. 99, no. 3, pp. 709–712, 2018.

[14] M. B. Laurens *et al.*, "A consultation on the optimization of controlled human malaria infection by mosquito bite for evaluation of candidate malaria vaccines," *Vaccine*, vol. 30, no. 36, pp. 5302–5304, 2012.

[15] J. E. Epstein *et al.*, "Safety and clinical outcome of experimental challenge of human volunteers with *Plasmodium falciparum*-infected mosquitoes: An update," *J. Infect. Dis.*, vol. 196, no. 1, pp. 145–154, Jul. 2007.

[16] United States Food and Drug Administration, "Guidance for industry toxicity grading scale for healthy adult and adolescent volunteers enrolled in preventive vaccine clinical trials," 2007. Accessed: Oct. 3, 2021, [Online]. Available: http://www.fda.gov/cber/guidelines.htm

[17] B. Mordmüller *et al.*, "Direct venous inoculation of *Plasmodium falciparum* sporozoites for controlled human malaria infection: A dose-finding trial in two centres," *Malar. J.*, vol. 14, pp. 117, Mar. 2015.

[18] M. Sklar *et al.*, "A three-antigen *Plasmodium falciparum* DNA prime—Adenovirus boost malaria vaccine regimen is superior to a two-antigen regimen and protects against controlled human malaria infection in healthy malaria-naïve adults," *PLoS One*, vol. 16, no. 9, Sep. 2021, Art. no. e0256980.

[19] J. E. Sasaki, D. John, and P. S. Freedson, "Validation and comparison of Actigraph activity monitors," *J. Sci. Med. Sport.*, vol. 14, no. 5, pp. 411–416, Sep. 2011.

[20] J. Karjalainen and M. Viitasalo, "Fever and cardiac rhythm," *Arch. Intern. Med.*, vol. 146, no. 6, pp. 1169–1171, Jun. 1986.

[21] J. Radin *et al.*, "Assessment of prolonged physiological and behavioral changes associated with COVID-19 infection," *JAMA Netw. Open*, vol. 4, no. 7, Jul. 2021, Art. no. e2115959.