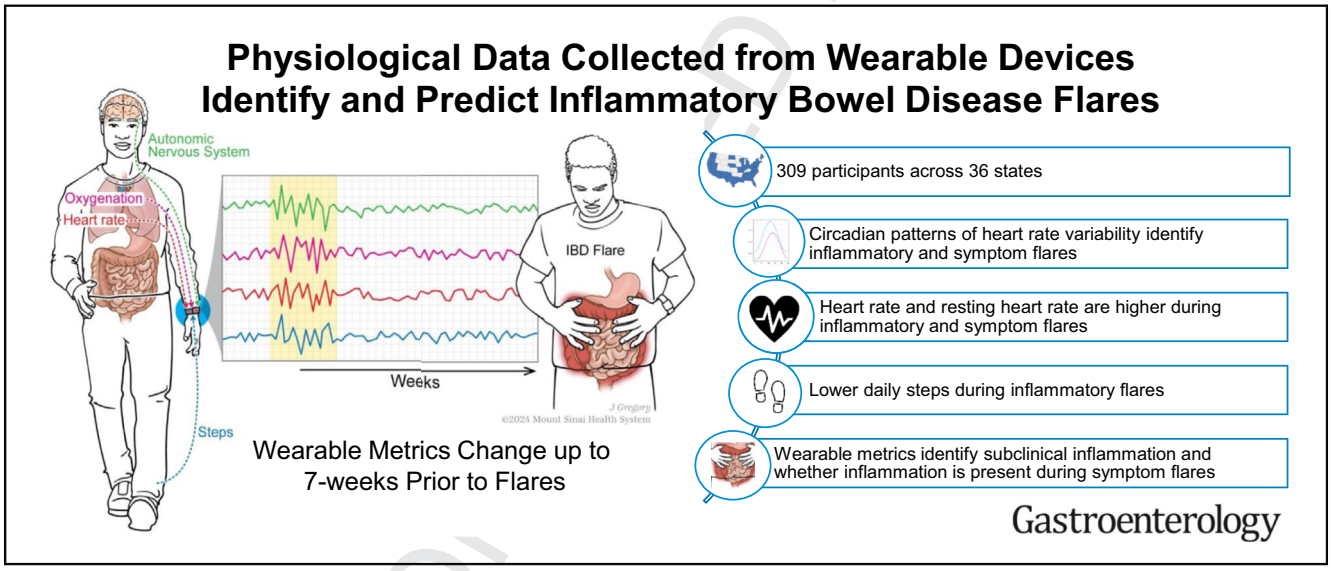
Gastroenterology 2025;■:1–13

Physiological Data Collected From Wearable Devices Identify and Predict Inflammatory Bowel Disease Flares

Q31 Robert P. Hirten,[1,2,3] Matteo Danieletto,[2,3] Milagros Sanchez-Mayor,[4] Jessica K. Whang,[1] Kyung Won Lee,[4] Kyle Landell,[2,3] Micol Zweig,[2,3] Drew Helmus,[1] Thomas J. Fuchs,[2,3] Zahi A. Fayad,[5] Girish N. Nadkarni,[6,7] Laurie Keefer,[1] Mayte Suarez-Farinas,[4] and Bruce E. Sands[1]

[1]The Dr Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, New York; [2]The Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, New York; [3]Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, New York; [4]Department of Population Health Science and Policy, Center for Biostatistics, Icahn School of Medicine at Mount Sinai, New York, New York; [5]Biomedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, New York; [6]Division of Data-Driven and Digital Medicine, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, New York; and [7]The Charles Bronfman Department of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York



BACKGROUND & AIMS: Wearable devices capture physiological signals noninvasively and passively. Many of these parameters have been linked to inflammatory bowel disease (IBD) activity. We evaluated the associative ability of several physiological metrics with IBD flares and how they change before the development of flare. **METHODS:** Participants throughout the United States answered daily disease activity surveys and wore an Apple Watch (Apple), Fitbit (Google), or Q7 Oura Ring (Oura Health). These devices collected longitudinal heart rate (HR), resting heart rate (RHR), heart rate variability (HRV), steps, and oxygenation. C-reactive protein, erythrocyte sedimentation rate, and fecal calprotectin were collected as standard of care. Linear mixed-effect models were implemented to analyze HR, RHR, steps, and oxygenation, and cosinor mixed-effect models were applied to HRV circadian features. Mixed-effect logistic regression was used to determine the predictive ability of physiological metrics. **RESULTS:** Three hundred and nine participants were enrolled across 36 states. Circadian patterns of HRV differed significantly between periods of inflammatory flare and remission and symptomatic flare and remission. Marginal means for HR and RHR were higher during periods of inflammatory flare and symptomatic flare. There were fewer daily steps during inflammatory flares. HRV, HR, and RHR differentiated whether participants with symptoms had inflammation. HRV, HR, RHR, steps, and oxygenation were significantly altered up to 7 weeks before inflammatory and symptomatic flares. **CONCLUSIONS:** Longitudinally collected physiological metrics from wearable devices can identify and change before IBD flares, suggesting their feasibility to monitor and predict IBD activity.

Keywords: Wearable Devices; Inflammatory Bowel Disease; Prediction; Crohn's Disease; Ulcerative Colitis.

WHAT YOU NEED TO KNOW

BACKGROUND AND CONTEXT

Inflammatory bowel disease flares are challenging to predict and rely on cross-sectional, single time point assessments.

NEW FINDINGS

Wearable devices collect nearly continuous physiological metrics that can identify and predict inflammatory and symptomatic inflammatory bowel disease flares.

LIMITATIONS

Laboratory measurements of disease activity were collected as part of participant's standard of care management, limiting the ability to determine precise transitions points between uninflamed and inflamed states.

CLINICAL RESEARCH RELEVANCE

This is the first large study to demonstrate the physiological metrics collected from wearable devices can identify, differentiate, and predict inflammatory and symptomatic flares. This indicates the potential ability of wearable devices to be used in the monitoring of inflammatory conditions, such as inflammatory bowel disease.

BASIC RESEARCH RELEVANCE

The dense physiological assessments wearable devices collect can potentially be used to complement and extend physiological observations about inflammatory bowel disease.

Inflammatory bowel diseases (IBDs)—ulcerative colitis (UC) and Crohn's disease (CD)—are chronic relapsing and remitting inflammatory diseases of the gastrointestinal tract.[1,2] Their management is challenging due to unpredictable and frequent disease flares and frequent discordance of inflammation and symptoms.[3] Current modalities of disease monitoring have substantial limitations and rely on patient-reported symptoms and assessment of inflammation using blood or stool testing, imaging, or ileocolonoscopy. However, evaluations are limited to a single time point of assessment, are inconvenient, and can be invasive. New monitoring modalities are needed that can assess disease activity longitudinally, passively, and in real time.

Wearable devices are used by approximately 20% of individuals in the United States.[4] They can measure physiological metrics in a near-continuous or continuous manner, including heart rate (HR), activity, resting heart rate (RHR), and heart rate variability (HRV). Wearable devices are an increasingly accepted tool for monitoring health and disease. They are frequently used in noninflammatory-based diseases for remote patient monitoring, allowing individuals to be monitored outside of the clinical setting, which has resulted in improved outcomes in multiple disease states.[5,6] There is increasing interest in monitoring chronic inflammatory diseases and prediction of flares and medication response. There are limited publications in this space due to the difficulty monitoring inflammatory-based conditions remotely via wearable technology. Individuals with IBD are interested in using wearable technology to manage their disease.[7] Yet, there are limited studies published that evaluate wearable technology in relation to IBD activity. One small study followed 37 patients with IBD after bowel surgery, finding that activity and sleep data from wearable biosensors predicted postoperative length of hospital stay.[8] Two studies evaluating a sweat-sensing wearable device further demonstrated that sweat-based assessments of immune and inflammatory markers correlate with serum measurements. These studies, although small, demonstrated the feasibility of passively monitoring markers of disease activity in IBD.[9,10]

Commonly used wearable devices can assess physiological signals, which are altered in chronic inflammatory diseases, thereby creating an opportunity for disease monitoring. Chronic inflammatory diseases, such as IBD, are associated with impaired autonomic nervous system (ANS) function. Both UC and CD have been found to have an impaired cardiovagal tone and relative parasympathetic suppression and sympathetic predominance.[11–13] ANS alterations are associated with changes in several physiological metrics that can be measured with wearable technology, providing an opportunity to monitor inflammatory diseases using these tools. An important parameter that can be measured by wearable technology is HRV. HRV is a measure of the small-time differences between each heartbeat. Changes in the beat-to-beat HR are controlled by both sympathetic (SNS) and parasympathetic nervous system (PNS) activity. HRV is an indirect measure of the ANS.[14] Higher HRV reflects a higher PNS tone, and lower HRV reflects higher SNS tone. Cross-sectional studies have demonstrated a decrease in HRV in IBD. In addition, they have shown that parasympathetic function displays a stronger inverse association with disease activity compared with sympathetic function supporting overall ANS dysfunction in IBD.[15] Similarly, HR is affected by changes in ANS function with increased sympathetic tone raising HR.[16]

To demonstrate the feasibility of wearable devices to monitor physiological metrics and their relationship to IBD disease activity, we performed a pilot study in which 15 participants with UC used a wearable patch to measure HRV over 9 months. We found that longitudinally collected increased SNS

Abbreviations used in this paper: **ANS, autonomic nervous system; AUC, area under the curve; AUPRC, area under precision-recall curve; BMI, body mass index; CD, Crohn's disease; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; FC, fecal calprotectin; HR, heart rate; HRV, heart rate variability; IBD, inflammatory bowel disease; MESOR, midline-estimating statistic of rhythm; PNS, parasympathetic nervous system; PRO, patient-reported outcomes; RMSSD, root mean square of successive differences between normal heartbeats; RHR, resting heart rate; SDNN, standard deviation of the inter-beat interval of normal sinus beats; SNS, sympathetic nervous system; SpO_2, oxygen saturation of arterial hemoglobin; UC, ulcerative colitis.** Q1 Q2

0016-5085/$36.00
https://doi.org/10.1053/j.gastro.2024.12.024

activity was associated with UC symptoms and elevated inflammatory markers.[17] Our current study built on this initial observation. To evaluate the ability of multiple physiological signals collected from wearable devices to identify IBD flares, we launched the IBD Forecast Study.

Materials and Methods

Study Population

The IBD Forecast Study is a prospective, exploratory, observational, cohort study enrolling individuals throughout the United States. Eligible participants had received a diagnosis of CD or UC, were 18 years or older, were on a medication to treat IBD, and were willing to use a wearable device. Participants were excluded if they were pregnant; had significant medical conditions, such as heart failure or another chronic disease; had a pacemaker or defibrillator; or used medications that impact HR or HRV, such as beta blockers, calcium channel blockers, or benzodiazepines.

Participants were nationally recruited via social media advertising, study flyers, a study website, and an e-mail about the study sent to individuals with IBD who visited the Inflammatory Bowel Disease Center at Mount Sinai Hospital, New York, NY. A participant could use their own Apple Watch (Apple), Fitbit (Google), or Oura Ring (Oura Health), or be provided with an Apple Watch or Fitbit for the duration of the study. No compensation was provided. Participants were able to contribute for as long as they liked. The study opened in December 2021 and closed in June 2023.

Study Procedures

Participants downloaded ehive, our custom digital research platform, to their smartphones; self-verified inclusion and exclusion criteria; and signed the electronic study consent.[18] Participants linked their wearable device or requested a wearable device from the study team, provided baseline demographic information, medical and surgical history, and IBD history through the app. Participants were recommended to use their wearable device for a minimum of 8 hours per day. Participants who used their wearable device or answered daily surveys fewer than 4 days per week were sent a smartphone push notification reminding them to participate.

Wearable Devices and Physiological Metrics

This study used 4 physiological metrics from the 3 device types (ie, Apple Watch, Fitbit, and Oura Ring). The metrics collected by all the devices included HRV, HR, RHR, and daily steps. These wearable devices are commercially available and contain photoplethysmography optical sensors. These sensors enable the assessment of capillary volume changes, determination of number of heart beats, and calculation of time between heart beats for HRV measurements.[19–21] HR is calculated as beats per minute throughout the day and night and RHR is defined as an estimation of a user's lowest HR during periods of rest. The Apple Watch calculates the SD of the inter-beat interval of normal sinus beats (SDNN), which is a time domain HRV metric that reflects both SNS and PNS activity. Lower values reflect increased sympathetic tone. The Fitbit and Oura Ring calculate HRV in the time domain metric root mean square of successive differences between normal heartbeats (RMSSD). The RMSSD is influenced more by the PNS compared with the SNS, with lower values reflecting increased sympathetic tone.[22] The Apple Watch Series 6 or later can measure the oxygen saturation of arterial hemoglobin (SpO_2) using conventional pulse oximetry methods. This feature has been validated in several studies.[23] See Supplementary Materials and Methods for further information.

Longitudinal Clinical and Laboratory Assessment

Clinical disease activity was assessed using baseline and daily UC- or CD-specific Patient Reported Outcome (PRO-2) scores. The PRO-2 is derived from the CD Activity Index and asks participants to rate their abdominal pain and stool frequency with the components weighted to create a cumulative score. The UC-specific PRO-2 is derived from the rectal bleeding and stool frequency sub-scores of the Mayo Clinic Score for UC. It consists of 2 questions that ask participants to rate their degree of rectal bleeding and stool frequency. C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and fecal calprotectin (FC) results were collected as part of each participant's standard of care management. Results were reported by participants in ehive. Laboratory values that were not reported were captured through the participants' electronic health record if he or she was followed at Mount Sinai Hospital.

Symptomatic and Inflammatory Flare Criteria

IBD flares were characterized as being consistent with inflammatory flares, denoted by elevated inflammatory biomarkers, or symptomatic flares, according to the presence of symptoms consistent with an IBD flare. In each timeframe, patients could meet the study definition of inflammatory flare either with or without symptomatic flare, symptomatic flare without inflammation, or remission. Inflammatory flare was defined by elevation of CRP, ESR, or FC values. An FC >150 μg/g was consistent with active inflammation.[24,25] CRP and ESR values were considered elevated if they were >5 mg/dL and >30 mm/h, respectively.[26] We assumed that the underlying inflammatory state extended for a period before and after the laboratory test is obtained. To account for this, and to avoid misclassification of inflammatory status around each laboratory test, CRP, ESR, and FC values were input in a window of ± 7 days. For the CD PRO-2, remission was defined as a PRO-2 score <8.[27,28] In UC, symptomatic remission was defined using a PRO-2 score ≤ 1 with a rectal bleeding score of 0 and stool frequency score ≤ 1.[29]

There is no consensus for the minimum number of daily symptom-based surveys denoting symptomatic activity required to classify a symptomatic flare. To evaluate potential criteria, we performed a sensitivity analysis to assess several definitions of symptomatic disease flare described in detail in Supplementary Figure 1. Based on this analysis, we limited our definition of symptomatic flare to having at least 4 surveys answered in each 7-day period, with at least 2 of these surveys meeting the criteria for symptomatic flare. This was consistent with our *a priori* definition of study compliance, which asked participants to answer a minimum of 4 surveys per 7-day period.

Statistical Analysis

Sex, age, device type, and body mass index (BMI) data were used as covariates in all statistical analyses. No association was

found between patients with reported or missing demographic information, supporting no departure from missing at random mechanism. As such, missing values were imputed based on the multivariate distribution of the observed values using the function are gImpute from Hmisc R package.[30] Longitudinal wearable measures were not imputed for this analysis. We use a mixed-model approach that produces robust estimates when missing values in the wearable data collection are missing at random. Factors leading to missing wearable data are dictated by technological and behavioral factors not easily determined. However, considering the large dataset, comparisons made within individuals, and models using wearable data collected when clinical information is recorded, the missing at random assumption is reasonably met.

Heart Rate Variability Analysis

Longitudinal changes in HRV have been found to have daily circadian patterns over a 24-hour period and can be modeled using a nonlinear Cosinor model approach.[31] Because our analysis evaluates the changes in the daily pattern as a function of flares in a longitudinal study, those changes were evaluated using Cosinor mixed-effects models. The analyses were carried out using the CosinorRmixedeffects R package,[32] which is publicly available and was developed by our team. Five hundred bootstraps were applied. This nonlinear model is fully described by Hou et al.[32] This approach models the HRV circadian rhythm over 24 hours. This pattern can be described using the following circadian parameters: midline-estimating statistic of rhythm (MESOR), a rhythm-adjusted mean; amplitude, a measure of one-half the extent of variation within a day; and acrophase, a measure of the time that overall high values occur each day.[33] The model included inflammatory and symptomatic flare status, as well as age, sex, and BMI as covariates. Device type was included as a fixed effect. Further details are available in the Supplementary Materials and Methods.

Heart Rate, Resting Heart Rate, Oxygen Saturation of Arterial Hemoglobin, and Steps Analysis

Linear mixed-effect models were fitted to analyze the physiological parameters of HR, RHR, and steps for all devices, and SpO_2 solely for the Apple Watch. Metrics were analyzed on an hourly basis (ie, HRV, SpO_2, and HR) and daily basis (ie, RHR and steps). Models include fixed effects for covariates of age, sex, and BMI, along with inflammatory and symptomatic flare predictors (and the interaction for Figure 2), as well as a random intercept for each participant for per-day models, and a random intercept nested by day for all per-hour models. These models were fitted using the lme function from the nlme R package.[34] Furthermore, heterogeneous variances within flare and remission stage were incorporated into our modeling with the final model defined in terms of likelihood ratio test and Akaike information criterion, as described above. Marginal means with 95% CIs were estimated using the emmeans package,[35] as well as testing the hypothesis of interests, namely, differences between flare stages through contrasts. To evaluate the effect of the presence or absence of nocturnal or daytime wearable data on outcomes, we performed a sensitivity analysis. RHR collected at night vs RHR collected over 24 hours did not result in significant differences (results not presented) in outcomes, demonstrating the negligible effect of the presence or absence of nocturnal data.

Physiological Parameters as an Early Signal of Flares

Mixed-effect logistic regression was used to evaluate the temporal association of various physiological parameters, or their combination, in determining the probability of experiencing flare-ups on the same day or before the flare-up periods occurs. Specifically, for each flare definition, we evaluated the changes in the physiological parameter taken at 7, 14, 21, 28, 35, 42, and 49 days before inflammatory or symptomatic flares. To evaluate models' performance, specificity, sensitivity, accuracy, precision, recall, area under precision-recall curve (AUPRC), area under the curve (AUC), and the F1-score were estimated. Further details are available in the Supplementary Materials and Methods.

Results

Cohort Characteristics

A total of 309 participants consented and contributed both wearable device and survey data across 36 states within the United States. Participants mean (SD) age was 39.9 (14.4) years and 208 (67.3%) were female. One hundred ninety-six participants had CD and 113 had UC. The mean (SD) duration of follow-up was 213 (156) days. Two hundred fifty-five participants wore an Apple Watch, 53 participants wore a Fitbit, and 16 participants wore an Oura Ring; some participants used 2 types of devices during the study period (Table 1). In total, 120 devices were loaned to participants (107 Apple Watches and 13 Fitbits). There were 152 days in inflammatory flare and 999.7 seven-day periods defined as symptomatic flares. Symptomatic and inflammatory trajectories of each participant are illustrated in Supplementary Figure 2. Statistics for mean wearable device use are presented in Supplementary Table 1.

Identification of Inflammatory Flares

HRV differentiated inflammatory flares compared with inflammatory remission, with significant differences in the circadian pattern of HRV measurements observed across device types (Figure 1*A*). The MESOR of the circadian pattern of SDNN was lower (39.26 ms; 95% CI, 36.69–42.10 ms) during an inflammatory flare compared with periods of inflammatory remission (41.06 ms; 95% CI, 38.65–43.80 ms; $P < .0001$). There were no differences in the acrophase ($P = .40$), amplitude ($P = .87$), or time to peak ($P = .40$) of the circadian measurements. There was a significant difference in the MESOR (38.31 ms; 95% CI, 32.63–44.58 ms; $P = .03$) of the circadian pattern of RMSSD in participants in an inflammatory flare compared with those without inflammation (MESOR: 35.87 ms; 95% CI, 29.88–42.41 ms). There was no difference in the acrophase ($P = .07$), amplitude ($P = .15$), or peak time ($P = .07$).

Physiological data differentiated inflammatory periods from noninflammatory periods in participants with IBD

Table 1. Demographic Characteristics for Study Participants

Characteristic	Participants with CD (n = 196)	Participants with UC (n = 113)	Overall cohort (n = 309)
Age, *y*, mean (SD)	39.6 (14.3)	40.5 (14.7)	39.9 (14.4)
Missing values, n (%)	0 (0)	1 (0.9)	1 (0.3)
Sex, female, n (%)	134 (68.4)	74 (65.5)	208 (67.3)
BMI, mean (SD)	25.7 (5.51)	25.7 (5.63)	25.7 (5.54)
Missing values, n (%)	10 (5.1)	9 (8.0)	19 (6.1)
Race, n (%)			
Asian	7 (3.6)	7 (6.2)	14 (4.5)
Black	3 (1.5)	3 (2.7)	6 (1.9)
Native American	2 (1.0)	1 (0.9)	3 (1.0)
White	178 (90.8)	100 (88.5)	278 (90.0)
Missing values	6 (3.1)	2 (1.8)	8 (2.6)
Ethnicity, n (%)			
Hispanic	11 (5.6)	7 (6.2)	18 (5.8)
Not Hispanic	179 (91.3)	101 (89.4)	280 (90.6)
Missing values	6 (3.1)	5 (4.4)	11 (3.6)
Smoking status, n (%)			
Current	6 (3.1)	3 (2.7)	9 (2.9)
Never	146 (74.5)	91 (80.5)	237 (76.7)
Past smoker	42 (21.4)	19 (16.8)	61 (19.7)
Missing values	2 (1.0)	0 (0)	2 (0.6)
Reported medical conditions, n (%)			
Asthma	27 (13.8)	8 (7.1)	35 (11.3)
History of cancer	9 (4.6)	3 (2.7)	12 (3.9)
Chronic kidney disease	2 (1.0)	0 (0)	2 (0.6)
Chronic lung disease	2 (1.0)	1 (0.9)	3 (0.9)
Diabetes mellitus	3 (1.5)	1 (0.9)	4 (1.3)
Hypertension	9 (4.6)	5 (4.4)	14 (4.5)
No medical conditions	141 (71.9)	91 (80.5)	232 (75.1)
History of IBD-related surgery, n (%)			
No	95 (48.5)	109 (96.5)	204 (66.0)
Yes	97 (49.5)	4 (3.5)	101 (32.7)
Missing values	4 (2.0)	0 (0)	4 (1.3)
IBD medications use at enrollment, n (%)			
Mesalamines	15 (7.7)	35 (31.0)	50 (16.2)
Corticosteroids	16 (8.2)	19 (16.8)	35 (11.3)
Biologic agents	129 (65.8)	37 (32.7)	166 (53.7)
Small molecules	0 (0)	6 (5.3)	6 (1.9)
Immune modulators	17 (8.7)	5 (4.4)	22 (7.1)
No medication reported	8 (4.1)	5 (4.4)	13 (4.2)
CD location, n (%)			
L1	89 (45.4)	—	—
L2	25 (12.8)	—	—
L3	74 (37.8)	—	—
L4	47 (24.0)	—	—
CD behavior, n (%)			
B1	39 (19.9)	—	—
B2	102 (52.0)	—	—
B3	55 (28.1)	—	—
Perianal modifier	59 (30.1)	—	—
UC location, n (%)			
Left-sided disease	—	39 (34.5)	—
Extensive colitis	—	38 (33.6)	—
Proctitis	—	11 (9.7)	—

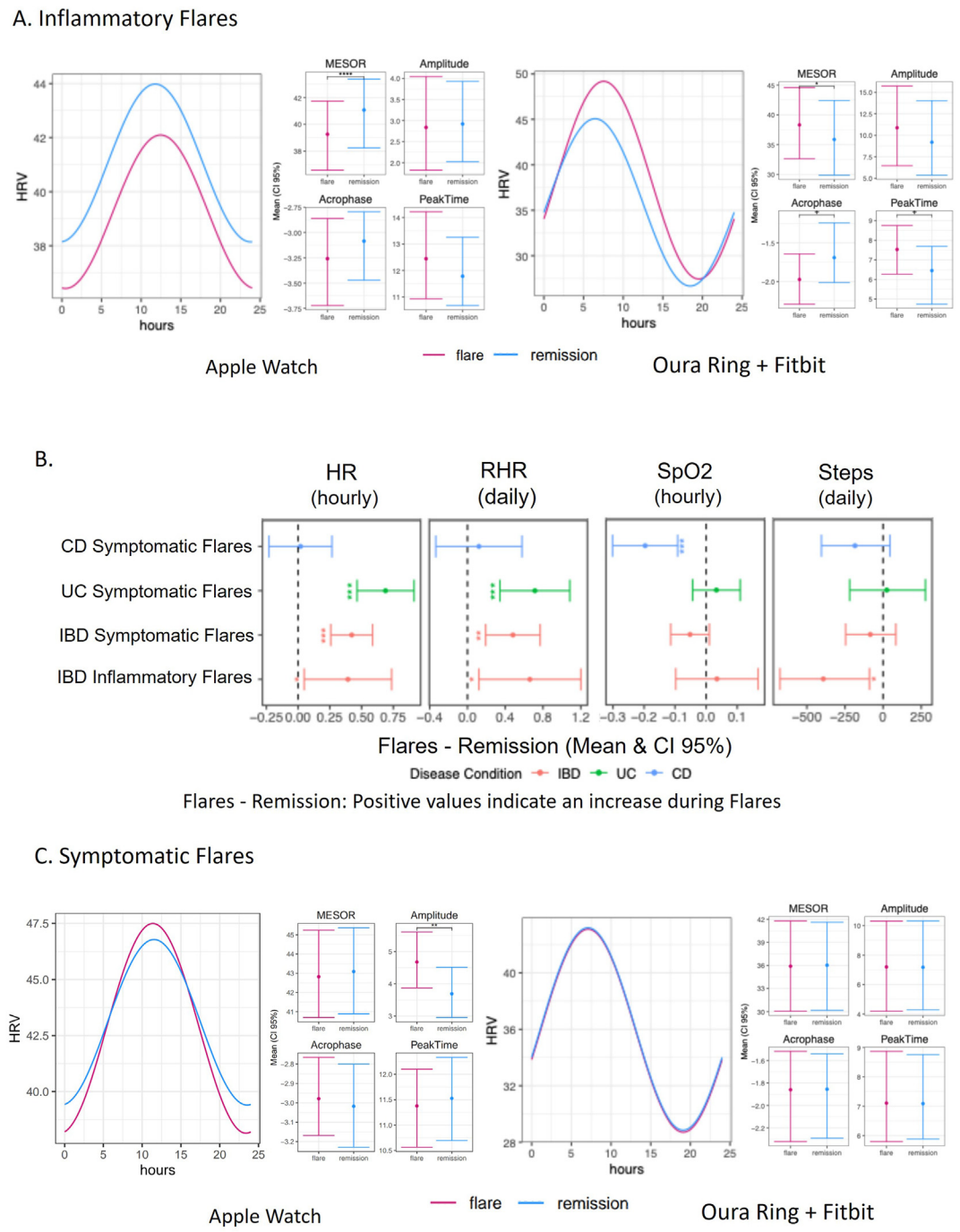


Figure 1. The relationship between physiological metrics collected from wearable devices and inflammatory and symptomatic flares. (*A*) HRV circadian patterns during periods of inflammatory flares compared with periods of inflammatory remission with means and 95% CI for HRV measures. (*B*) The HR, RHR, SpO_2, and daily steps during periods of symptomatic and inflammatory flares compared with periods of remission. (*C*) HRV circadian patterns during periods of symptomatic flares compared with periods of symptomatic remission with estimated marginal means and 95% CIs for HRV parameters. *Stars* represent the comparison between flare and remission states. $^{+}P \leq .1$; $^{*}P \leq .05$; $^{**}P \leq .01$; $^{***}P \leq .001$; $^{****}P \leq .0001$.

(Figure 1*B*). Mean (SEM) HR and RHR were higher during inflammatory flares (HR: 79.31 [0.92] beats/min; RHR: 65.30 [0.80] beats/min) and daily steps were fewer (steps: 5082 [0.02]) compared with periods of inflammatory remission (HR: 78.92 [0.92] beats/min; $P = .03$; RHR: 64.64 [0.78] beats/min; $P = .02$; steps: 5507 [0.02]; $P = .01$). Mean (SEM) SpO_2 was not different during inflammatory flares (96.52% [0.14%]) compared with periods of inflammatory remission (96.49% [0.13%]; $P = .61$).

Identification of Clinical Flares

Inflammatory bowel disease. Physiological metrics differentiate symptomatic flares from symptomatic

remission in those with IBD. In participants with IBD, including both CD and UC, there were significant differences in the circadian pattern of HRV measurements across devices and measurement types. The amplitude of the circadian pattern of SDNN was higher (4.68 ms; 95% CI, 3.87–5.60 ms) during symptomatic flares compared with periods without symptoms (3.69 ms; 95% CI, 2.96–4.51 ms; $P = .006$) (Figure 1*C*). There was no difference in the acrophase ($P = .66$), MESOR ($P = .26$), or peak time ($P = .66$) of the circadian pattern of HRV. There were no differences observed in the circadian pattern of RMSSD during symptomatic flares compared with periods without symptoms (MESOR, $P = .81$; amplitude, $P = .96$; acrophase, $P = .96$; peak time, $P = .96$).

Mean (SEM) HR and RHR were higher during symptomatic flares (HR: 78.32 [0.68] beats/min; RHR: 64.60 [0.62] beats/min) compared with periods of symptomatic remission (HR: 77.89 [0.68] beats/min; $P < .0001$; RHR: 64.11 [0.61]; $P = .0009$) (Figure 1*B*). Mean (SEM) daily steps and SpO_2 were not different during symptomatic flares (steps: 5564 [0.02]; SpO_2: 96.50% [0.11%]) compared with periods in symptomatic remission (steps: 5649 [0.02]; $P = .32$; SpO_2: 96.55% [0.10]; $P = .10$).

Crohn's disease. In participants with CD, there were no differences in the circadian pattern of SDNN (MESOR, $P = .83$; amplitude, $P = .44$; acrophase, $P = .40$; peak time, $P = .40$) or RMSSD (MESOR, $P = .06$; amplitude, $P = .38$; acrophase, $P = .65$; peak time, $P = .65$) during a symptomatic flares compared with symptomatic remission. Mean (SEM) daily steps were not different between symptomatic flares of CD compared with periods of symptomatic remission (5193 [0.02] vs 5385 [0.02], respectively; $P = .11$). There was no significant difference in mean (SEM) HR (78.10 [0.92] beats/min vs 78.08 [0.91] beats/min, respectively; $P = .88$) and RHR (64.69 [0.85] vs 64.55 [0.84]; $P = 0.58$) between the 2 states. However, SpO_2 was significantly lower during symptomatic flares compared with periods of symptomatic remission (96.44% [0.14%] vs 96.63% [0.14%], respectively; $P = .0002$) between the 2 states (Figure 1*B*).

Ulcerative colitis. In participants with UC, the circadian pattern of SDNN significantly differed between periods of symptomatic flare compared with periods of remission. There was a significant difference in the amplitude ($P = .002$) of the circadian pattern of SDNN between periods of symptomatic flares (5.30 ms; 95% CI, 4.07–6.67 ms) and symptomatic remission (3.76 ms; 95% CI, 2.81–4.97) ms), although there were no differences seen in the other circadian features of HRV (MESOR, $P = .24$; acrophase, $P = .89$; peak time, $P = .89$). Differences in the circadian patterns of RMSSD were able to differentiate periods of symptomatic flare (MESOR, 37.34 ms; 95% CI, 28.71–46.06 ms) compared with symptomatic remission (34.52 ms; 95% CI, 25.90–43.16 ms; $P < .0001$). Daily steps did not differ during symptomatic flares of UC compared with periods of symptomatic remission (mean [SEM] steps: 6103 [0.02] vs 6079 [0.02], respectively; $P = .85$). There was a significant increase in mean (SEM) HR (78.89 [1.04] beats/min vs 78.20 [1.04] beats/min, respectively; $P < .0001$) and RHR (64.53 [0.92] beats/min vs 63.81 [0.91] beats/min; $P = .0001$) during symptomatic flares compared with periods of symptomatic remission. There was no difference in mean (SEM) SpO_2 (96.56% [0.16%] vs 96.53% [0.16%], respectively; $P = .40$) between periods of symptomatic flare and remission (Figure 1*B*).

Interactions Between Symptoms and Inflammation

We sought to determine whether physiological metrics collected from wearable devices can differentiate the presence and absence of symptoms and underlying inflammation. The MESOR ($P < .0001$), acrophase ($P < .0001$), and peak time ($P < .0001$) of SDNN significantly differed between periods of symptomatic and inflammatory flare compared with periods with symptomatic flare but no inflammation (Supplementary Table 2), thus identifying whether inflammation is present or absent during symptomatic periods. In addition, there was a significant difference in the MESOR ($P < .0001$), acrophase ($P = .03$), and peak time ($P = .03$) of the circadian pattern of SDNN during periods with and without symptomatic flares in the presence of underlying inflammation. There were no significant differences seen in the circadian pattern of SDNN during periods in symptomatic remission and underlying inflammation compared with periods in both symptomatic and inflammatory remission. There were no differences in the MESOR, amplitude, acrophase, or peak time of the circadian pattern of SDNN in participants with a symptomatic flare compared with those in symptomatic remission, if there was no underlying inflammatory flare ($P > .05$) (Figure 2*A*).

Similar findings were observed when evaluating other physiological metrics collected from the wearable devices. Periods with symptomatic flares and inflammatory flares compared with periods with symptomatic flares but in inflammatory remission were differentiated by mean (SEM) RHR (67.20 [1.09] beats/min vs 65.28 [0.96] beats/min, respectively; $P = .006$) and HR (81.08 [1.02] vs 78.43 [1.00] beats/min, respectively; $P < .001$), but not SpO_2 (96.35% [0.20%] vs 96.13% [0.17%], respectively; $P = .07$). Comparing periods of symptomatic remission and inflammatory flare to periods with no symptoms and no inflammation, there was no difference in mean (SEM) RHR and HR (64.60 [0.96] beats/min vs 64.40 [0.92] beats/min, respectively, $P = .62$; 77.57 [1.00] vs 77.64 [0.99] beats/min, respectively; $P = .82$). However, mean (SEM) SpO_2 was able to differentiate these 2 states (96.26% [0.19] vs 96.52% [0.17%], respectively; $P = .04$). Mean (SEM) RHR, HR, and SpO_2 were able to differentiate periods of symptomatic flare compared with symptomatic remission if there was no inflammation present (65.28 [0.96] beats/min vs 64.40 [0.92] beats/min; $P = .03$; 78.43 [1.00] beats/min vs 77.64 [0.99] beats/min; $P = .003$; 96.13% [0.17%] vs 96.52% [0.17%]; $P < .0001$, respectively) (Figure 2*B*).

Change Preceding Inflammatory and Symptom Flares

Physiological metrics were evaluated for their changes 7, 14, 21, 28, 35, 42, and 49 days before flares of IBD

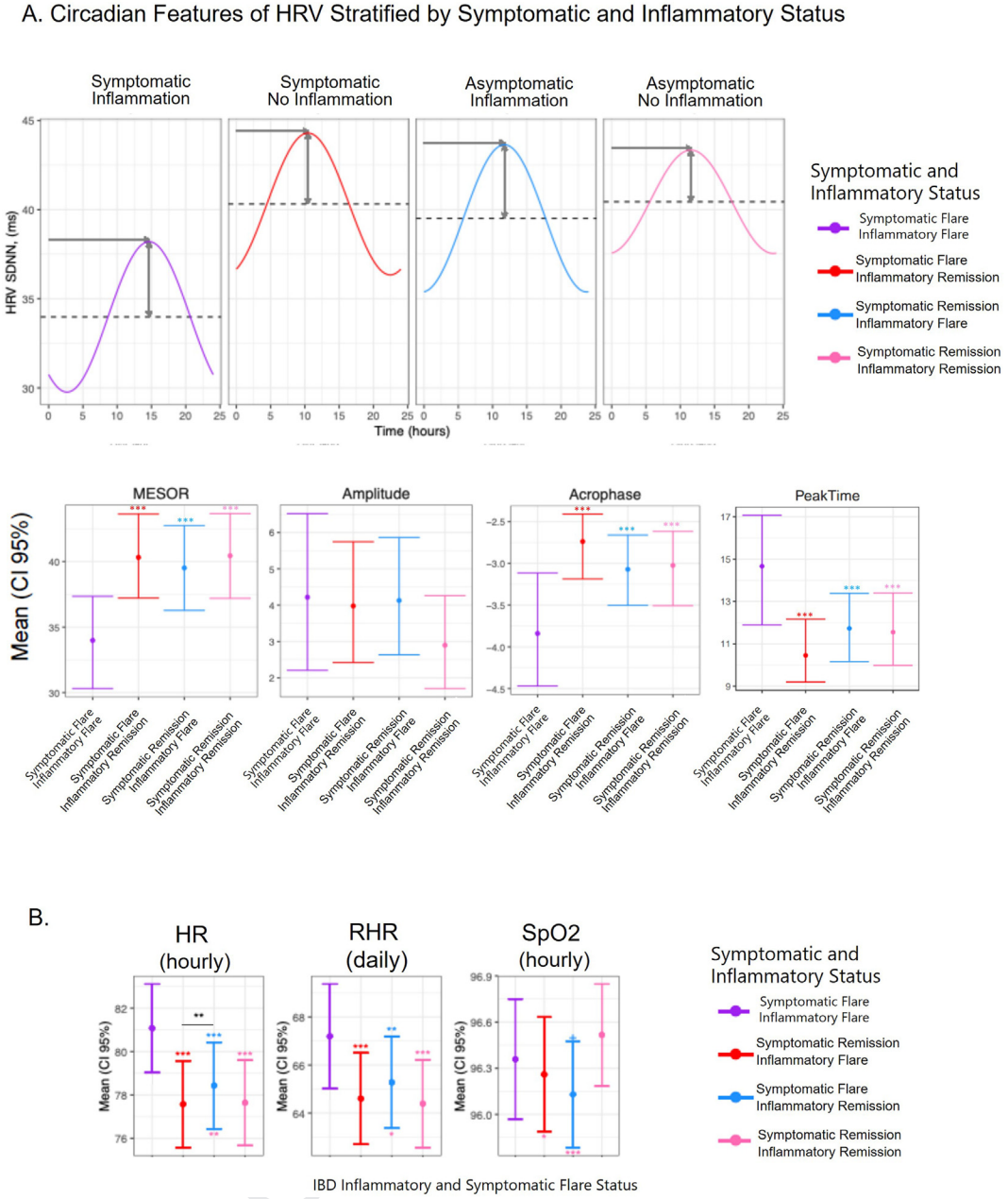


Figure 2. Analysis of physiological metrics by both inflammatory and symptom status in participants using the Apple Watch (HRV, SpO_2) and all devices (HR, RHR). (*A*) The mean HRV circadian patterns for each interaction state is presented and the estimated marginal means and 95% CIs for HRV metrics measured during each state are shown. (*B*) HR, RHR, and SpO_2 metrics stratified based on each inflammatory and symptom status with estimated marginal means and 95% CIs for HRV measures. *Stars* atop the upper 95% CI are comparisons with the symptomatic and inflammatory flare (*purple*) group. *Stars* under the lower 95% CI are comparisons with symptomatic and inflammatory remission (*pink*) group. Other comparisons are indicated in *black*. $^{+}P \leq .1$; $^{*}P \leq .05$; $^{**}P \leq .01$; $^{***}P \leq .001$; $^{****}P \leq .0001$

(Figure 3*A*). Analysis was limited to Apple Watch–derived metrics to enable direct comparison of the changes in measured metrics before flares. HRV (SDNN), HR, RHR, steps and SpO_2, as well as models combining these metrics, maintained high F1 scores throughout the 49-day period before inflammatory and symptomatic flares (Figure 3*B*). F1 scores balance precision and recall and are optimal for evaluating imbalanced datasets. When evaluating the physiological parameters that were analyzed on a per-day basis (ie, HR, steps, and RHR), a model including all metrics maintained an F1 score of 0.88 at 49 days preceding identification of inflammatory flares (49 days before flare: AUC, 0.98; 95% CI, 0.97–0.99; F1, 0.88; AUPRC, 0.55; sensitivity, 0.92; specificity, 0.94) (Supplementary Figure 3*A* and *B*, Supplementary Table 3). Similar results were found when evaluating physiological parameters analyzed on a per-hour basis (ie, HRV, HR, and SpO_2). A model including all of these metrics had an F1 score of 0.90 at 49 days preceding

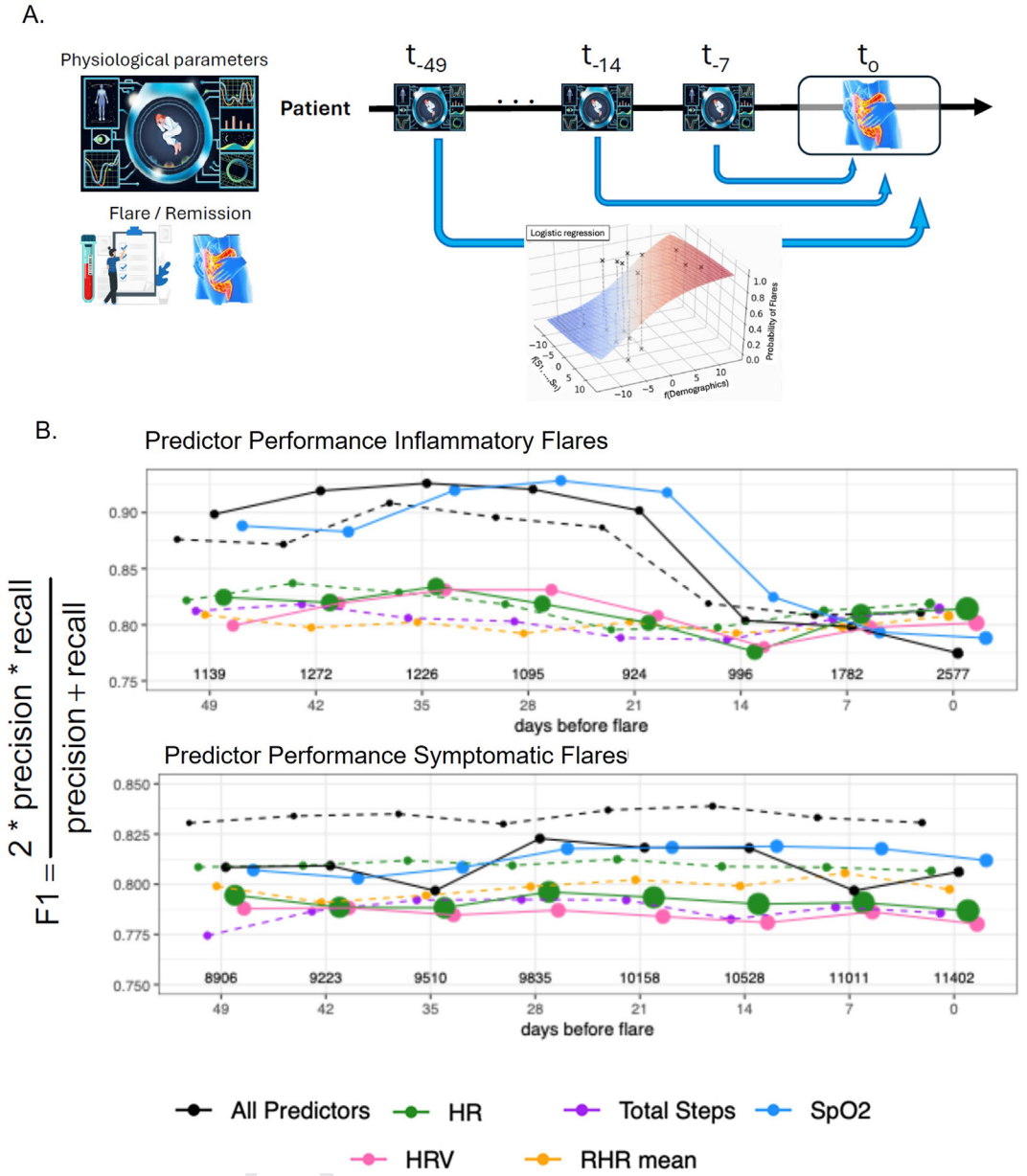


Figure 3. (*A*) Association of early physiological metric (taken 7, 14, 21, 28, 35, 42, and 49 days before flare periods) with subsequent inflammatory and symptomatic flares up to 7 weeks before the events were assessed. (*B*) The F1 value summarizes the performance of the model that includes individual physiological metrics or a combination of them at each time point. The F1 score balances precision and recall and is optimal for evaluating imbalanced datasets. The F-score ranges from 0.0 to 1.0, representing perfect precision and recall. The number of physiological data points analyzed at each time point is listed at each time interval. *Solid lines* represent physiological metrics analyzed on a per-hour basis. *Dashed lines* represent metrics analyzed on a per-day basis.

identification of inflammatory flares (49 days before flare; AUC, 0.99; 95% CI, 0.99–0.99; F1, 0.90; AUPRC, 0.60; sensitivity, 0.91; specificity, 0.97) (Supplementary Table 4). The F1 scores for symptomatic flares were similarly high and largely stable over the 49 days before flare (Supplementary Figure 3*A* and *B*). A model including the parameters measured on a per-day basis (ie, HR, steps, and RHR) had an F1 score of 0.83 at 49 days preceding the flare event (49 days before flare; AUC, 0.96; 95% CI, 0.96–0.97; F1, 0.83; AUPRC, 0.46; sensitivity, 0.89; specificity, 0.87) (Supplementary Table 5). Similar results were found for the parameters measured on a per-hour basis (ie, HRV, HR, and SpO_2). A model including these parameters had an F1 score of 0.81 at 49 days before the flare event (49 days before flare; AUC, 0.96; 95% CI, 0.96–0.96; F1 0.81; AUPRC, 0.49; sensitivity 0.90; specificity 0.86) (Supplementary Table 6).

Discussion

The IBD Forecast study demonstrates the feasibility of identifying and predicting IBD flares using noninvasive commonly used wearable technologies. We observed that

the physiological metrics collected from wearable devices are altered up to 7 weeks before the development of inflammatory and symptomatic flare periods. This is the first study, to our knowledge, that demonstrates that wearable devices can identify the worsening of a chronic inflammatory disease, as well as differentiate whether underlying inflammation is present in symptomatic and asymptomatic individuals.

IBD monitoring relies on cross-sectional and often invasive or inconvenient means to assess disease activity. This results in long periods during which no disease assessments occur. Furthermore, there is no convenient means to predict the development of inflammation or disease flares. The advancement of digital technologies provides an opportunity for frequent, passive, and real-time assessments of an individual's physiological status, filling in the data gaps generated by traditional disease assessments. The ability of wearable devices to passively collect data, coupled with their popularity, has generated increasing interest in their use to monitor health and disease.

Optical and actigraphy sensors on wearable devices can frequently measure several important physiological metrics. ANS dysfunction, characterized by an uncoupling of the ANS and hypothalamic–pituitary–adrenal axis, has been found in individuals with IBD, with greater degrees of dysregulation seen in individuals with active disease compared with remission.[36] Indirect measures of ANS activity, specifically HRV, have been cross-sectionally correlated to IBD activity in several studies. IBD is associated with overall decreases in HRV indices, with this inverse relationship strongest for the PNS.[15] Our group furthered these observations by demonstrating a longitudinal relationship between inflammation and SNS tone, and an inverse relationship between inflammation and parasympathetic tone.[17] HR is similarly affected by the ANS through innervation of the sinoatrial node of the heart, with the SNS increasing.[16] Thus, physiological metrics that are measured by wearable devices are altered by ANS dysfunction and the SNS activation seen with active inflammation in IBD. The ability of such metrics to identify and predict underlying inflammatory events has been demonstrated in studies using wearable devices for the identification and prediction of COVID-19 infections. Such studies demonstrated significant increases in RHR and mean HR and altered HRV circadian patterns during and before an infection.[31,37–39]

We similarly observed significant changes in the physiological metrics measured by wearable devices during IBD flares. There were significant changes in HRV circadian patterns, which differentiate periods of inflammation from remission. The MESOR of the circadian pattern for SSDN was lower during flares, reflecting increased sympathetic tone. Furthermore, when evaluating the RMSSD metric, we noted other changes in the HRV circadian pattern, demonstrating an overall alteration of ANS activity. An increase in mean HR and RHR were observed during periods of inflammation. Similar changes in all 3 physiological metrics were observed during symptomatic flares.[27–29] Interestingly, we found that daily steps were lower during inflammatory flares, however, we did not find a difference in steps with symptomatic flares. This observation is in line with 1 unpublished study demonstrating lower steps the week before inflammatory marker elevation.[40] Although this warrants further evaluation, an interpretation of daily steps can be challenging, as physical activity measurements are prone to bias due to device nonadherence and are not measuring a physiological parameter.

Beyond the identification of flares, we found that physiological metrics collected from wearable devices can identify inflammatory and clinical flares up to 7 weeks early. This observation highlights that measurable physiological changes occur before the development of flare events. Similar observations have been made with FC, which can rise 6 months before the development of clinical flares.[41] This supports our observation that there can be a lag time between physiological status and flare. In the case of wearable-based metrics, this change preceding flares may similarly be secondary to the development of subclinical inflammation, which alters ANS activity, and the parameters measured by wearable devices that it impacts. One limitation of our study design is that regular inflammatory assessments are not obtained, precluding determination of when individuals transition from remission to flare. To address this limitation, our group is evaluating transitions between remission and flare states and their impact on wearable-based metrics, to further explore this finding.

The physiological source for many wearable-based biomarkers creates an opportunity to better characterize the disease status of individuals with IBD. There is discordance between symptoms and inflammation in IBD.[42,43] We observed that HRV, HR, and RHR were able to differentiate the presence or absence of inflammation in symptomatic individuals. This analysis demonstrates that in IBD, inflammation is a primary driver of the changes in wearable measured physiological metrics. However, our observation that HR and RHR could differentiate symptomatic from asymptomatic periods, even in the absence of inflammation, shows that symptoms alone have some impact on physiological parameters. Further evaluation of this observation is needed to understand the degree and contribution of symptoms and inflammation on wearable-derived metrics. Interestingly, we observed that differences in systemic oxygen levels were predictive of inflammatory and clinical disease flares. Although the intestinal epithelium is normally in a state of relative hypoxia, which is exacerbated during active inflammation, there have not been descriptions of systemic oxygen levels measured during disease flares.[44] The pathophysiology explaining this observation is not clear, and may be secondary to alterations of the oxygen saturation curve, which result in slight changes in pulse oximetry measurements, which can be secondary to carbon dioxide concentrations, acid–base status, and temperature.[45]

Our findings demonstrate the possibility of digitally collected physiological metrics to serve as a novel biomarker of disease activity. Although there is no gold standard for detecting IBD flares, monitoring using wearable devices may potentially be better, given their continuous and physiological grounding. Although the

physiological changes we observed during periods of flare differed significantly compared with periods of remission, the absolute differences in these values were small. These small differences, however, are often sufficient for individual flare detection via machine learning and deep learning algorithms. This has been performed in other conditions with similar findings, resulting in a clinically relevant outcome.[37,46] Other areas warranting further exploration include the utility of wearable biomarkers to identify and predict IBD medication response,[47] as well as their ability to serve as future end points in clinical trials.

There are several limitations to our study. First, the physiological metrics measured by the wearable devices are not specific and can be impacted by other factors, including intercurrent conditions. To mitigate this with regard to HRV, we controlled for the covariates of BMI, age, and sex in our analysis. However, this is an important area that requires further study and validation. In addition, inflammatory assessments were collected as part of standard of care evaluations. This limits our ability to determine when an individual transitions from inflammatory remission to inflammatory flare. To address this limitation, we imputed blood and stool results ±7 days around each collection time point, to account for the fact that, in IBD, inflammatory markers are elevated for extended periods of time around flares.[48] However, because we do not know when individuals transition from remission to inflammation, there is the potential for the misclassification of disease activity periods. Similarly, several potential definitions can be used to determine periods of symptomatic flare compared with symptomatic remission. This includes assumptions around the number of daily surveys that need to be answered per 7-day period, and the number of surveys consistent with symptomatic activity in a 7-day period needed to classify that period as "flare." Although we explored this in a sensitivity analysis, there are no well-defined definitions in the literature, therefore, leading to the potential to misclassify symptomatic periods. In addition, there is a limitation in our mixed-effect logistic regression models used to predict flare-ups from physiological data. Overall, the study dataset is imbalanced, with a high prevalence of both inflammatory and symptomatic remission states compared with flare events. Furthermore, many flares occur within the same individuals. This low variability within patients can result in some biasing of the model's outcome and AUC results. However, we settled on this approach, as the dataset contains repeated measures, with individuals contributing multiple outcomes (ie, flares). An additional limitation is that the impact of medication is not controlled in the analyses. There is significant variation in medication type, dose, frequency, and the timing of medication changes. There are variations in how the medications overlap with provided symptomatic and inflammatory assessments. This makes it challenging to control for medications in the analysis. Lastly, digital study cohorts have been found to differ in comparison with nondigital study cohorts, including in composition and adherence to chronic medical management. Although this is a limitation to all digital studies, it is important to recognize the potential for bias or nongeneralizability of results that this can introduce.

We found that physiological metrics collected longitudinally from wearable devices can identify and change before the development of inflammatory and symptomatic disease flares. Furthermore, physiological metrics can differentiate whether there is underlying inflammation present during symptomatic flares. These findings support the further evaluation of wearable devices in the monitoring of IBD.

Supplementary Material

Note: To access the supplementary material accompanying this article, visit the online version of *Gastroenterology* at www.gastrojournal.org, and at http://doi.org/10.1053/j.gastro.2024.12.024.

References

1. Torres J, Mehandru S, Colombel JF, et al. Crohn's disease. Lancet 2017;389:1741–1755.
2. Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. Lancet 2017;389:1756–1770.
3. Colombel JF, Keir ME, Scherl A, et al. Discrepancies between patient-reported outcomes, and endoscopic and histological appearance in UC. Gut 2017;66:2063–2068.
4. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Available at: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/. Accessed XXXX.
5. Teo E, Hassan N, Tam W, et al. Effectiveness of continuous glucose monitoring in maintaining glycaemic control among people with type 1 diabetes mellitus: a systematic review of randomised controlled trials and meta-analysis. Diabetologia 2022;65:604–619.
6. Ehringer DS, Mughmaw TE, Albers RC. Use of remote patient monitoring kits to reduce hospitalization and mortality rates for patients with heart failure. Am J Health Syst Pharm 2024;81(Suppl 1):S15–S20.
7. Hirten RP, Stanley S, Danieletto M, et al. Wearable devices are well accepted by patients in the study and management of inflammatory bowel disease: a survey study. Dig Dis Sci 2021;66:1836–1844.
8. Yi Y, Sossenheimer PH, Erondu AI, et al. Using wearable biosensors to predict length of stay for patients with IBD after bowel surgery. Dig Dis Sci 2022;67:844–853.
9. Hirten RP, Lin KC, Whang J, et al. Longitudinal assessment of sweat-based TNF-alpha in inflammatory bowel disease using a wearable device. Sci Rep 2024;14:2833.
10. Hirten RP, Lin KC, Whang J, et al. Longitudinal monitoring of IL-6 and CRP in inflammatory bowel disease using IBD-AWARE. Biosens Bioelectron X 2024;16:100435.
11. Lindgren S, Stewenius J, Sjölund K, et al. Autonomic vagal nerve dysfunction in patients with ulcerative colitis. Scand J Gastroenterol 1993;28:638–642.
12. Ganguli SC, Kamath MV, Redmond K, et al. A comparison of autonomic function in patients with inflammatory bowel disease and in healthy controls. Neurogastroenterol Motil 2007;19:961–967.

13. Aghdasi-Bornaun H, Kutluk G, Keskindemirci G, et al. Evaluation of autonomic nervous system functions in frame of heart rate variability in children with inflammatory bowel disease in remission. Turk J Pediatr 2018; 60:407–414.
14. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 1996;17:354–381.
15. Kim KN, Yao Y, Ju SY. Heart rate variability and inflammatory bowel disease in humans: a systematic review and meta-analysis. Medicine (Baltimore) 2020;99: e23430.
16. Gordan R, Gwathmey JK, Xie LH. Autonomic and endocrine control of cardiovascular function. World J Cardiol 2015;7:204–214.
17. Hirten RP, Danieletto M, Scheel R, et al. Longitudinal autonomic nervous system measures correlate with stress and ulcerative colitis disease activity and predict flare. Inflamm Bowel Dis 2021;27:1576–1584.
18. Hirten RP, Danieletto M, Landell K, et al. Development of the ehive digital health app: protocol for a centralized research platform. JMIR Res Protoc 2023;12:e49204.
19. Apple. Monitor Your Heart Rate with Apple Watch. Available at: https://support.apple.com/en-us/HT204666. Accessed October 22, 2020.
20. McMullen M. How Fitbit trackers monitor heart rate. Available at: https://healthsolutions.fitbit.com/blog/how-do-fitbit-trackers-monitor-heart-rate/. Accessed December 16, 2023.
21. Oura. Live heart rate. Available at: https://support.ouraring.com/hc/en-us/articles/4410651298963-Live-Heart-Rate. Accessed December 16, 2023.
22. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017; 5:258.
23. Blood oxygen app on Apple Watch. Apple Updated October 2022.
24. Sandborn WJ, Panés J, Zhang H, et al. Correlation between concentrations of fecal calprotectin and outcomes of patients with ulcerative colitis in a phase 2 trial. Gastroenterology 2016;150:96–102.
25. Gisbert JP, Bermejo F, Pérez-Calle JL, et al. Fecal calprotectin and lactoferrin for the prediction of inflammatory bowel disease relapse. Inflamm Bowel Dis 2009; 15:1190–1198.
26. Croft A, Lord A, Radford-Smith G. Markers of systemic inflammation in acute attacks of ulcerative colitis: what level of c-reactive protein constitutes severe colitis? J Crohns Colitis 2022;16:1089–1096.
27. Peyrin-Biroulet L, Panés J, Sandborn WJ, et al. Defining disease severity in inflammatory bowel diseases: current and future directions. Clin Gastroenterol Hepatol 2016; 14:348–354.e17.
28. Khanna R, Zou G, D'Haens G, et al. A retrospective analysis: the development of patient reported outcome measures for the assessment of Crohn's disease activity. Aliment Pharmacol Ther 2015;41:77–86.
29. Jairath V, Khanna R, Zou GY, et al. Development of interim patient-reported outcome measures for the assessment of ulcerative colitis disease activity in clinical trials. Aliment Pharmacol Ther 2015;42:1200–1210.
30. Harrell F Jr. Hmisc: Harrell Miscellaneous. R package, version 5.1-2. Available at: https://hbiostat.org/R/Hmisc/. Accessed XXXX. Q22
31. 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.
32. Hou R, Tomalin LE, Suárez-Fariñas M. cosinoRmixedeffects: an R package for mixed-effects cosinor models. BMC Bioinform 2021;22:553.
33. Hirten RP, Danieletto M, Tomalin L, et al. Factors associated with longitudinal psychological and physiological stress in health care workers during the COVID-19 pandemic: observational study using Apple Watch data. J Med Internet Res 2021;23:e31295.
34. Pinheiro J, Bates D. R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package, version 3.1-162. Available at: https://cran.r-project.org/web/packages/nlme/index.html. Accessed XXXX. Q23
35. Lenth RV. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package, version 1.10.1. Available at: https://CRAN.R-project.org/package=emmeans. Accessed XXXX. Q24
36. Pellissier S, Dantzer C, Mondillon L, et al. Relationship between vagal tone, cortisol, TNF-alpha, epinephrine and negative affects in Crohn's disease and irritable bowel syndrome. PLoS One 2014;9:e105328.
37. Hirten RP, Tomalin L, Danieletto M, et al. Evaluation of a machine learning approach utilizing wearable data for prediction of SARS-CoV-2 infection in healthcare workers. JAMIA Open 2022;5:ooac041.
38. 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.
39. Mishra T, Wang M, Metwally AA, et al. Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 2020;4:1208–1220.
40. Sossenheimer P, Yvellez OV, Andersen MJ, et al. 539 - Wearable devices can predict disease activity in inflammatory bowel disease patients. Gastroenterology 2019; 156(Suppl 1):S-111.
41. Kostas A, Siakavellas SI, Kosmidis C, et al. Fecal calprotectin measurement is a marker of short-term clinical outcome and presence of mucosal healing in patients with inflammatory bowel disease. World J Gastroenterol 2017;23:7387–7396.
42. Ma C, Sandborn WJ, D'Haens GR, et al. Discordance between patient-reported outcomes and mucosal inflammation in patients with mild to moderate ulcerative colitis. Clin Gastroenterol Hepatol 2020; 18:1760–1768.e1.
43. Vasant DH, Ford AC. Functional gastrointestinal disorders in inflammatory bowel disease: time for a paradigm shift? World J Gastroenterol 2020;26:3712–3719.

44. Cummins EP, Crean D. Hypoxia and inflammatory bowel disease. Microbes Infect 2017;19:210–221.
45. StatPearls. 2024.
46. Hirten RP, Suprun M, Danieletto M, et al. A machine learning approach to determine resilience utilizing wearable device data: analysis of an observational cohort. JAMIA Open 2023;6:ooad029.
47. Vasudevan A, Gibson PR, van Langenberg DR. Time to clinical response and remission for therapeutics in inflammatory bowel diseases: what should the clinician expect, what should patients be told? World J Gastroenterol 2017;23:6385–6402.
48. Kennedy NA, Jones GR, Plevris N, et al. Association between level of fecal calprotectin and progression of Crohn's disease. Clin Gastroenterol Hepatol 2019; 17:2269–2276.e4.

Received May 29, 2024. Accepted December 24, 2024.

Correspondence
Address correspondence to: Robert P. Hirten, MD, The Dr Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, 1468 Madison Avenue, Annenberg Building, Room 5-12, New York, New York 10029. e-mail: robert.hirten@mountsinai.org.

CrediT Authorship Contributions
Robert P. Hirten, MD (Conceptualization: Lead; Funding acquisition: Lead; Methodology: Lead; Project administration: Lead; Writing – original draft: Lead; Writing – review & editing: Lead)
Matteo Danieletto, PhD (Data curation: Equal; Project administration: Supporting; Software: Lead; Writing – review & editing: Supporting)
Milagros Sanchez-Mayor, PhD (Formal analysis: Equal; Writing – original draft: Supporting; Writing – review & editing: Supporting)
Jessica K. Whang, MS (Project administration: Supporting; Writing – review & editing: Supporting)
Kyung Won Lee, PhD (Formal analysis: Supporting)
Kyle Landell, BS (Software: Supporting; Writing – review & editing: Supporting)
Micol Zweig, MS (Project administration: Supporting; Writing – review & editing: Supporting)
Drew Helmus, MS (Project administration: Supporting; Writing – review & editing: Supporting)
Thomas J. Fuchs, DrSc (Project administration: Supporting; Supervision: Supporting; Writing – review & editing: Supporting)
Zahi A. Fayad, PhD (Supervision: Supporting; Writing – review & editing: Supporting)
Girish N. Nadkarni, MD (Supervision: Supporting; Writing – review & editing: Supporting)
Laurie Keefer, PhD (Supervision: Supporting; Writing – review & editing: Supporting)
Mayte Suarez-Farinas, PhD (Formal analysis: Lead; Writing – original draft: Supporting; Writing – review & editing: Supporting)
Bruce E. Sands, MD (Conceptualization: Supporting; Funding acquisition: Supporting; Methodology: Supporting; Writing – review & editing: Supporting)

Conflicts of interest
The authors disclose the following: Robert P. Hirten reports consulting for Bristol Meyers Squibb. Laurie Keefer reports consulting for Dr Reddy's, Coprata Health, Trellus Health, Pfizer, Ardelyx, and AbbVie and equity ownership in Trellus Health. Zahi A. Fayad reports consulting for Rockley Photonics. Bruce E. Sands reports consulting fees from AbbVie, Abivax, Adiso Therapeutics, Agomab, Alimentiv, Amgen, AnaptysBio, Arena Pharmaceuticals, Artugen Therapeutics, AstraZeneca, Biora Therapeutics, Boehringer Ingelheim, Boston Pharmaceuticals, Bristol Myers Squibb, Calibr, Celgene, Celltrion, ClostraBio, Enthera, Equillium, Evommune, Ferring, Fresenius Kabi, Fzata, Galapagos, Genentech (Roche), Gilead Sciences, GlaxoSmithKline, Gossamer Bio, Index Pharmaceuticals, Innovation Pharmaceuticals, Janssen, Kaleido, Kallyope, Lilly, Merck, Mobius Care, Morphic Therapeutics, MRM Health, Nexus Therapeutics, Nimbus Discovery, Odyssey Therapeutics, Palisade Bio, Pfizer, Progenity, Prometheus Biosciences, Prometheus Laboratories, Protagonist Therapeutics, Q32 Bio, Rasayana Therapeutics, Sun Pharma, Surrozen, Takeda, Target RWE, Teva, Theravance Biopharma, TLL Pharmaceutical, TR1X, Union Therapeutics and Ventyx Biosciences; speaking fees from AbbVie, Abivax, Bristol Myers Squibb, Janssen, Lilly, Pfizer, and Takeda; research grants from Bristol Myers Squibb, Janssen, Pfizer, and Takeda; stock and stock options from Ventyx Biopharma; and other support from Abivax, Bristol Myers Squibb, Lilly, Janssen, Pfizer, and Takeda. The remaining authors disclose no conflicts.

Funding
Support for this study was provided by National Institute of Diabetes and Digestive and Kidney Diseases grant K23DK129835 to Robert P. Hirten. The authors would like to thank the generous support of Ms. Jenny Steingart.

Data Availability
Data collected as part of this study will be made available on reasonable request. Analytic methods and study materials will not be made available to other researchers beyond what is described in the current study.

Supplementary Materials and Methods

Wearable Devices and Physiological Metrics

All of the devices (ie, Apple Watch, Fitbit, and Oura Ring) contain 3-axis accelerometer signals that track motion, such as the number of steps per day. In addition, the 3 devices can calculate HRV. They can record time series peaks correlating with each heartbeat via their photoplethysmography optical sensor and are used to determine inter-beat intervals and, therefore, HRV. The Apple Watch specifically calculates each measurement of SDNN during a 60-second window throughout each 24-hour period, and Fitbit and Oura ring capture HRV readings every 5 minutes during periods of sleep. Our team acquired Fitbit-derived HRV data every 5 minutes during periods of sleep. Regarding oxygen assessment by the Apple Watch, the red to infrared modulation ratio is determined through the optical sensor, which correlates with the color of the underlying skin's arterial blood and is translated into the SpO_2 percentage.

Quality-control procedures were applied to each physiological variable, removing values outside expected physiological ranges, and assigned as a missing value. Limits outside the following ranges for each variable were assigned as missing values: SpO_2, 85%–100%; HR, 30–220 beats/min; RHR, 30–150 beats/min; steps, <0 steps/d; and HRV (SDNN, RMSSD), 5–300 ms.

Cosinor Models

HRV, among other physiological processes, follows a circadian pattern.[e1] Many wearable devices, including the Apple Watch, Fitbit, and Oura Ring, do not collect data continuously over 24-hour periods. HRV, as well as other physiological metrics, are collected sparsely and in a nonuniform manner. Therefore, traditional evaluations of HRV, such as calculation of its mean or range of values, fail to account for the circadian nature of the measurements and the dynamic changes in this pattern that occur in response to physiological events, such as flare. In addition, given the sparse and nonuniform measurement of HRV by wearable devices, such evaluations would introduce bias. Therefore, techniques that can model such datasets are needed for wearable-based physiological metrics. Daily circadian rhythms can be modeled using Cosinor methods. Cosinor methods have frequently been applied to evaluate HRV.[e2-e4] To account for longitudinal changes in circadian patterns, our team extended these Cosinor models to include a mixed-effect model framework.[32]

As described in the Materials and Methods section of the main text, the nonlinear Cosinor model can be used to describe circadian features, including the adjusted mean (MESOR) of the circadian pattern, the maximum change or height of the circadian pattern (MESOR), and the time the maximum amplitude or peak is achieved (acrophase).

Cosinor mixed-effect models can be described through the following linear model:

$$y_i(t) = M + a_0 c_i + (\beta + a_1 c_i)x_i(t) + (\gamma + a_2 c_i)z_i(t) + \theta_i w_i(\mathrm{t}) + e(t) \quad (1)$$

where $y_i(t)$ is the vector of hourly (t) observations every day related to HRV, M is the midline statistic of rhythm or overall rhythm-adjusted mean (MESOR), β and γ are the vectors of fixed effects related to nonlinear parameters: amplitude (A), representing the maximum change from the MESOR within a day, and the acrophase (ϕ) representing the time the amplitude is reached (peak $= -\boldsymbol{\phi} * 24 / 2\boldsymbol{\pi}$). Those parameters are linearly represented as $\beta = A\,cos(\phi)$ and $\gamma = A\,sin(\phi)$. a_0, a_1, and a_2 are fixed effects associated with the vector of covariates c, which included the inflammatory/symptomatic flare status as well as age, sex, and BMI as covariates. In addition, when modeling data from multiple devices, the type of device was included as a fixed effect. θ is the vector of random effects and e are the model residuals. All models considered a random intercept (MESOR) across participants, and we evaluated whether the additional random effects for β and γ (associated with the $x[t]$ and $z[t]$ components) improved model fitting. Models with different random effects were compared using the likelihood ratio test, where whether the test yielded significance, the model exhibiting a lower Akaike information criterion was selected. Otherwise, if the likelihood ratio test yielded no significance, the most parsimonious model would be chosen.[31]

Physiological Parameters as an Early Signal of Flares

The mixed-effect logistic regression was evaluated through the following model:.

$$logit(P(Flare_{t+i} = 1|X_t)) = \beta_0 + \beta_1 X_{it} + u_i Z_i + e_{it} \quad (2)$$

where $Flare_{t+i}$ represents the occurrence of a flare on day $t + i$ with $i = 7, 14, 21, 28, 35, 42,$ and 49. β_0 and β_1 are the coefficients of the model, representing the intercept and the effect of the covariates and predictors, respectively, on the log-odds of the response and u_i represents the random effect across participants. In general, given the physiological measures on day t, the model fits the probability of experiencing a flare on day $t + i$. Models were fit using glmer function from lme4 package.[e5]

Supplementary References

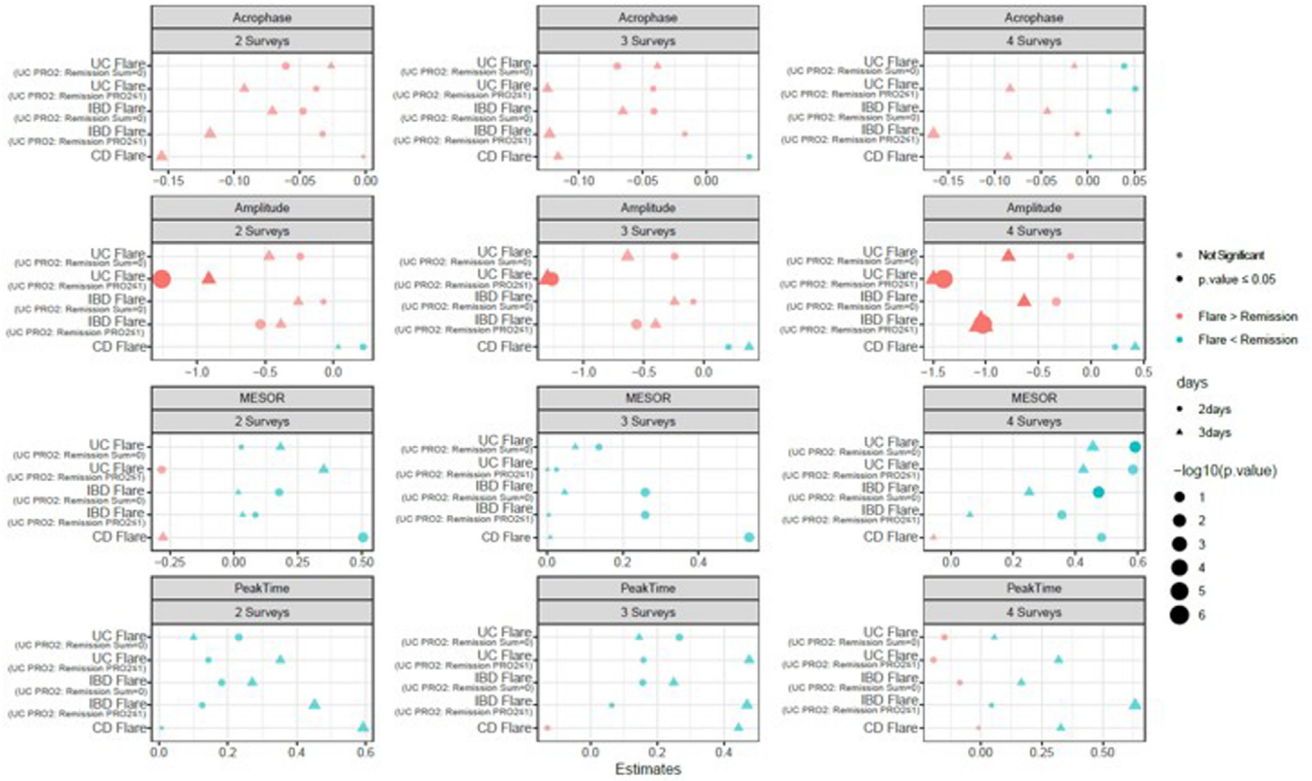
e1. Malpas SC, Purdie GL. Circadian variation of heart rate variability. Cardiovasc Res 1990;24:210–213.

e2. Tabata M, Takeshima T, Burioka N, et al. Cosinor analysis of heart rate variability in ambulatory migraineurs. Headache 2000;40:457–463.

e3. Yang Z, Liu H, Meng F, et al. The analysis of circadian rhythm of heart rate variability in patients with drug-resistant epilepsy. Epilepsy Res 2018;146:151–159.

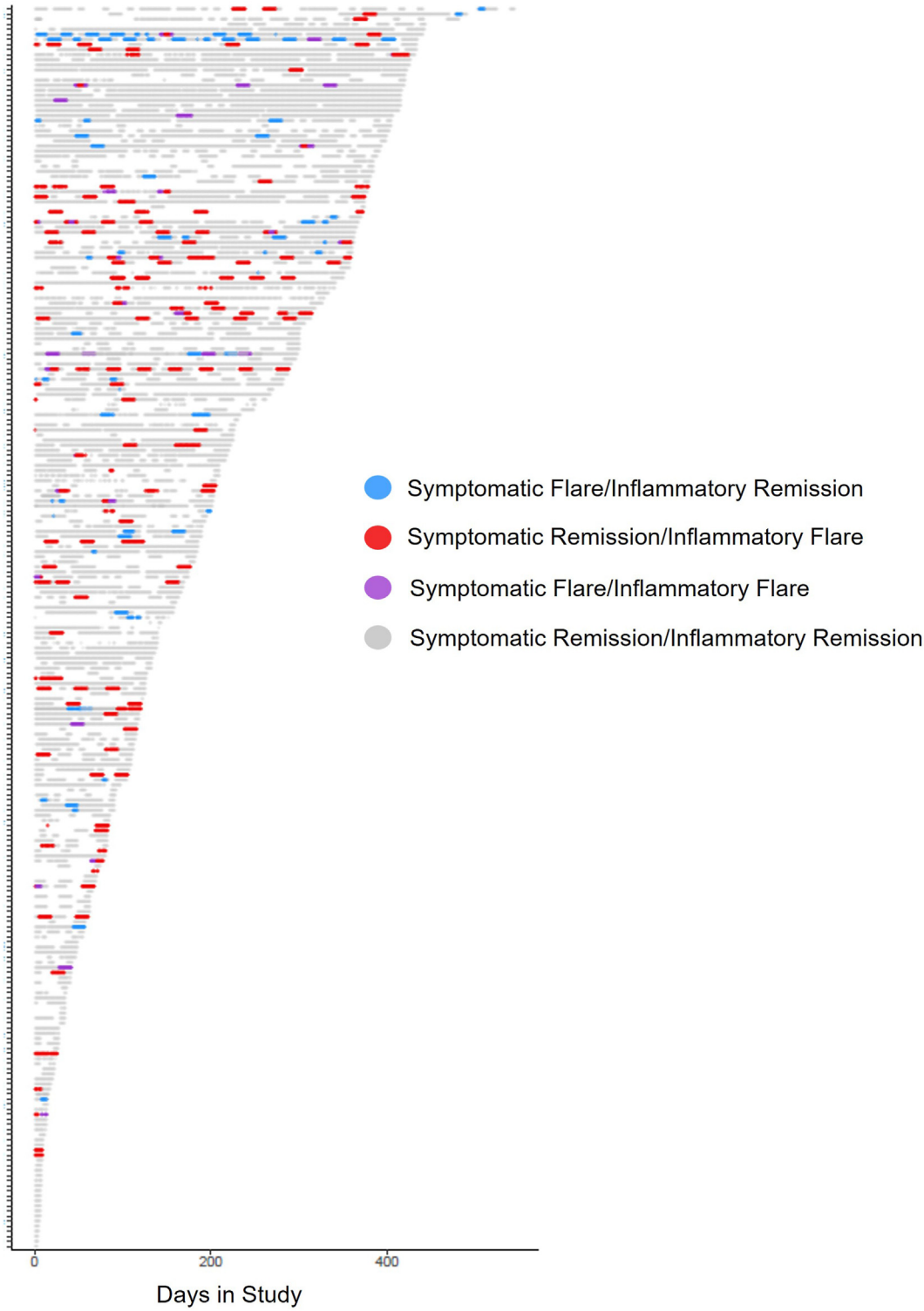
e4. Alkhodari M, Jelinek HF, Werghi N, et al. Investigating circadian heart rate variability in coronary artery disease patients with various degrees of left ventricle ejection fraction. Annu Int Conf IEEE Eng Med Biol Soc 2020; 2020:714–717.

e5. Bates D, Mächler M, Bolker B, et al. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1–48.

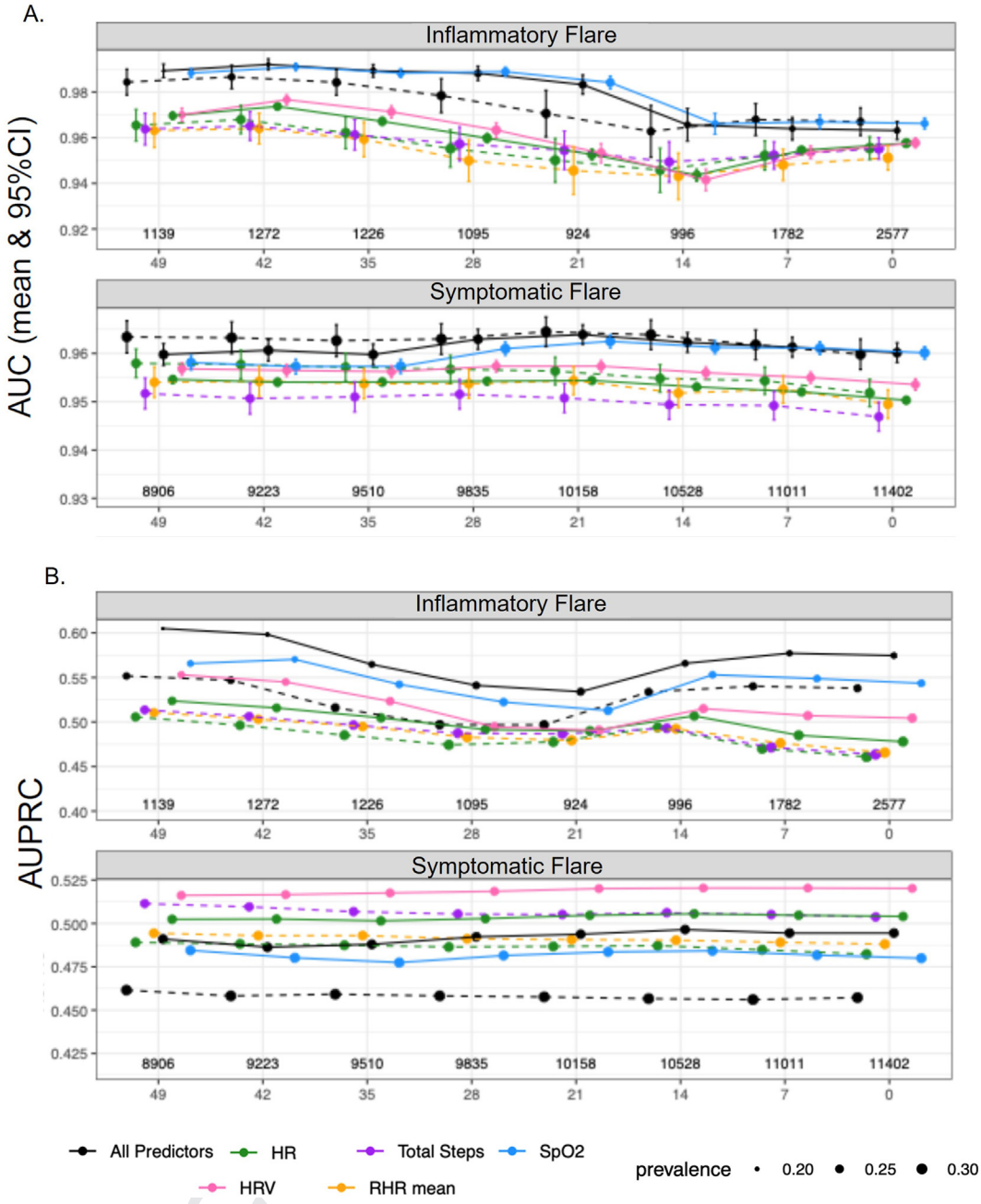


web 4C/FPO

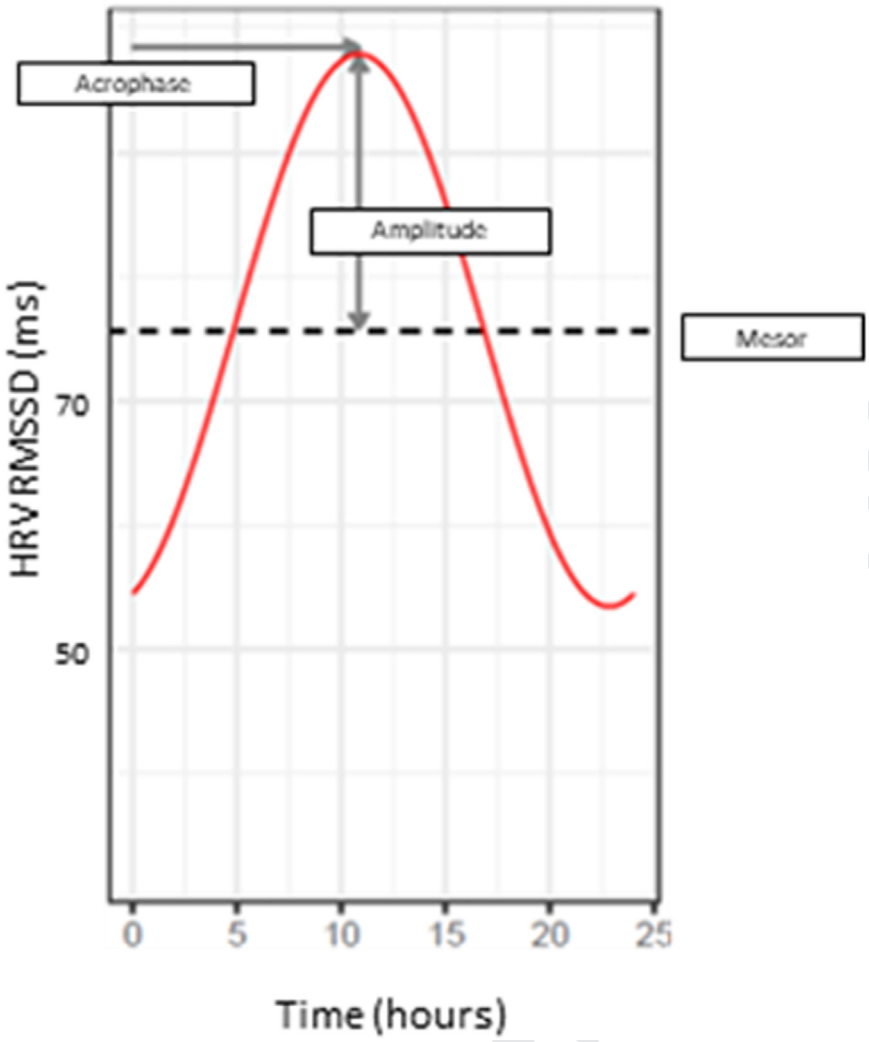
Supplementary Figure 1. The relationship between several symptomatic disease flare criteria was explored. This is given the absence of evidence supporting specific criteria for symptomatic flares derived from daily symptom assessments. The circadian features of HRV were estimated (as explained in Materials and Methods section) and stratified based on the number of questionnaires answered and the number of surveys consistent with a symptomatic flare, within a 7-day window. The number of surveys required to be completed in a 7-day window is listed in each box and ranges from 2 to 4 surveys. Within each of these periods, the *circles* denote having a minimum of 2 symptomatic days, and *triangles* denote a minimum of 3 symptomatic days defining flare, *red* (*blue*) color indicates an increase (decrease) in the physiological measures during flare compared with remission periods, the size of the symbols represents the level of significance (scaled as $-\log_{10}[P]$) with *darker colors* representing significant changes at .05 level. This was evaluated for UC (using 2 PRO-2 scoring cutoffs), CD, and IBD (both conditions combined). MESOR, midline statistic of rhythm; acrophase, representing the time the amplitude is reached (peak time = $-\phi^{*}24/2\pi$); amplitude, representing the maximum change from the MESOR within 1 day.



Supplementary Figure 2. Symptom and inflammatory activity of each participant over the study period. The *x-axis* denotes each day that a participant was in the study, with day 0 denoting the day of first data collection. Each *line* along the *y-axis* denotes an individual study participant. *Colors* denote the symptom and inflammatory activity.



Supplementary Figure 3. (*A*) The AUC with means and 95% CIs and (*B*) AUPRC represent the performance of models that incorporate individual physiological metrics or combinations thereof at each time point (taken at 7, 14, 21, 28, 35, 42, and 49 days before flare periods). Sample sizes used in each metric are indicated at each time point. *Solid lines* represent physiological metrics analyzed on a per-hour basis. *Dashed lines* represent metrics analyzed on a per-day basis.



Supplementary Figure 4. ■■■