npj | digital medicine

Article

Published in partnership with Seoul National University Bundang Hospital

https://doi.org/10.1038/s41746-026-02543-3

# Digital physiological biomarkers predict within-person symptom changes in complex chronic illness



Annie Aitken[1] ✉, Abbey Sawyer[2], Akiko Iwasaki[3,4], Harlan M. Krumholz[5,6], Rory Preston[7], Paul Calcraft[7], Harry Leeming[7], Jenna Tosto-Mancuso[2], Amy Proal[8], Michael A. Osborne[9] & David Putrino[2] ✉

Altered heart-rate variability (HRV) and resting heart rate (HR) are common in many complex chronic conditions. Mobile and wearable technologies now provide real-time, valid measurements of HRV and HR, advancing symptom monitoring and management. The current study integrates a 60-s morning PPG assessment with evening symptom severity reports, yielding a high-density mobile health dataset ($n = 4244$) with an average of 125 biometric observations per participant. We examined whether within-person fluctuations in HR, HRV, and respiratory rate predicted daily changes in crash, fatigue, and brain fog symptoms and secondarily evaluated model predictive performance. Model fit and variance explained were highest in models that included morning biometrics in addition to prior-day symptom reports and covariates. Within-person increases in HR and decreases in HRV in the morning were associated with worsening symptom reports in the evening. Walk-forward cross-validation showed a statistically significant improvement in model performance when morning biometrics were added to prior-day symptom reports (AUC = 0.82–0.85 vs. 0.73–0.83). These findings represent the prospective utility of mobile health tools for precision monitoring and prediction of real-time symptom exacerbations in complex chronic illness.

Complex chronic illnesses such as Long COVID (LC) and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) can significantly impact an individual's quality of life[1,2]. The fluctuating and often unpredictable nature of symptoms in these conditions further contributes to this decline. For example, approximately 85% of those affected by LC experience episodic symptoms that can rapidly fluctuate from periods of symptom stability to severe exacerbations, resulting in significant functional declines[3–5]. These periods of severe worsening symptoms are often colloquially called "crashing" or "flare-ups" by those living with complex chronic illnesses. However, biological drivers of these symptom exacerbations have not been fully elucidated, and often, people with complex chronic illnesses are unable to link a change in symptoms to a particular event. Overall, the inability of people with complex chronic illnesses to have measurable ways to predict periods of symptom exacerbation can lead to less effective disease management and, subsequently, reduced quality of life[5].

Recent work suggests that heart rate variability (HRV) and resting heart rate (HR) metrics may hold utility as potential biomarkers of complex chronic illness symptoms, such as LC, ME/CFS, or other energy-limiting conditions[6,7]. Resting HR, or the number of successive heartbeats per minute, and HRV, the variability of R to R intervals measured linearly or in time/frequency domains, have been identified as key metrics of overall health in healthy cohorts[8–10]. Further, HRV is shown to be a strong indicator of cardiovascular health, physiological responsiveness to stress, and autonomic function[9,11,12]. HRV, though generally non-specific, is influenced by multiple organ systems, including the cardiovascular, autonomic nervous, respiratory, and immune systems, making it a robust indicator of overall health. Given the diversity of organ systems and pathobiology implicated in the pathogenesis of complex chronic illnesses, serial evaluation of HRV and HR may represent a novel approach to identifying and predicting symptom exacerbations. For instance, it is theorized that LC involves disruption of autonomic nervous system balance. This autonomic disruption is thought

[1]New York University, New York, NY, USA. [2]Cohen Center for Recovery from Complex Chronic Illness, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Department of Immunobiology, Yale University, New Haven, CT, USA. [4]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [5]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA. [6]Department of Internal Medicine, Section of Cardiovascular Medicine, Yale School of Medicine, New Haven, CT, USA. [7]Visible Health Inc., Wilmington, DE, USA. [8]PolyBio Research Foundation, Medford, MA, USA. [9]Department of Engineering Science, University of Oxford, Oxford, UK. ✉e-mail: aitkenannie@gmail.com; david.putrino@mountsinai.org

to result from systemic inflammation and immune activation, including the release of pro-inflammatory cytokines that heighten sympathetic activity and suppress parasympathetic tone[13,14]. Indeed, researchers have found that individuals with LC exhibit atypical diurnal HRV patterns relative to controls[15], including reduced HRV and elevated resting HR[16,17].

In laboratory environments, HRV, HR, and respiratory rate (RR) are typically measured using electrocardiograms (ECG) and respiration belts. These tools have been clinically validated to provided research-reliable estimates of sympathetic (SNS) and parasympathetic nervous system (PNS) activity, but are generally limited to controlled lab or clinical settings[18–20]. Using ECG and respiration belts for managing health conditions is less practical in real-world settings due to their cumbersome nature and reliance on specialized equipment, making them difficult to implement. However, recent advances in mobile technologies have enabled naturalistic and user-friendly collection of cardiac and respiratory biomarkers of autonomic nervous system (ANS) control via photoplethysmography (PPG) sensors on smartphone cameras or wearable devices[21,22]. These advances in digital health allow for the acquisition of self-reported daily fluctuations in symptoms alongside biometric proxies of physiological fluctuation.

For example, biometrics and longitudinal symptom reporting have been pivotal in predicting symptom exacerbation in people with chronic diseases. One example is in people with chronic obstructive pulmonary disease (COPD). Using the myCOPD app, daily self-report data used in machine learning models could predict respiratory exacerbation in the subsequent days with moderate accuracy[23]. Other digital health applications for people with COPD such as COPDPredict™ accurately predicted respiratory exacerbations with a high degree of accuracy using both self-report and biometric inputs[24]. These models have the potential to influence early treatment of respiratory exacerbations, which is associated with improved clinical outcomes, including exacerbation recovery time. Similar platforms developed for cystic fibrosis have determined that the use of simple self-reporting platforms can lead to earlier detection of respiratory exacerbations and treatment in the form of oral antibiotics in adults[25] and pediatrics[26]. Finally, researchers have demonstrated that in individuals with Type 1 Diabetes, glucose fluctuations measured with continuous glucose monitors dynamically covary with changes in processing speed[27]. Similarly, the commercially available Accu-Chek® SmartGuide Predict app has shown utility in forecasting glucose levels in real time to reduce duration of hypoglycemia and hyperglycemia[28]. While the pathophysiology differs substantially, these studies raise questions about whether digital health tools could impact the course of complex chronic illnesses.

The Visible application is one such mobile application that has recently been developed that provides people with LC, ME/CFS, and other complex chronic conditions the ability to self-monitor symptoms and potentially relevant biometrics. The present study aims to conduct a high-level descriptive analysis of retrospective intensive longitudinal data from the Visible application. Using a hierarchical, stepwise multilevel modeling approach, we examined both within-person and between-person variation in daily biometric measures - specifically heart rate (HR), heart rate variability (HRV), and respiration rate (RR) - in relation to subsequent evening reports of crash, fatigue, and brain fog. In addition, we conducted an exploratory predictive modeling analysis to test whether models trained on past data could accurately generalize to future observations for the same individuals (within-person temporal generalization). Across these analyses, our primary goal was to quantify the incremental predictive value that biometric variables provide beyond that of prior-day symptom reports.

## Results

### Descriptive statistics

Data from $n = 4244$ Visible application users are presented. Table 1 presents descriptive statistics on the Visible dataset to reflect the breadth of the data used in the subsequent analyses. Descriptive graphics of the biometrics are presented in Fig. 1. Of the full sample, 3514 users tracked crashes, 3915 users tracked fatigue, and 3525 users tracked brain fog.

### Multilevel model regressions

We used hierarchical model building to evaluate the incremental contribution of lagged symptom reports and morning biometric indicators to the prediction of crash, fatigue, and brain fog outcomes (see Table 2). For all three outcomes, the inclusion of autoregressive lag terms (prior day symptom report) on top of covariates significantly improved model fit, as evidenced by increases in marginal $R^2$, substantial reductions in Akaike Information Criterion (AIC), and significant chi-squared model comparisons ($\Delta\chi^2 = 27{,}037.98$ for crash; $\Delta\chi^2 = 45{,}462.08$ for fatigue; $\Delta\chi^2 = 74{,}716.15$ for brain fog; all $p < 0.001$). Adding morning biometrics led to further improvements in model performance across all outcomes, supporting their added predictive value beyond covariates and autoregressive effects ($\Delta\chi^2 = 587.93$ for crash; $\Delta\chi^2 = 1259.07$ for fatigue; $\Delta\chi^2 = 402.50$ for brain fog; all $p < 0.001$). Full incremental model results are provided in the SI Tables 1–3.

We first used a logistic GLMER with random participant intercepts to model how fluctuation in daily morning biometrics predicted the likelihood of a reported crash that evening (Fig. 2a). On the within-person level, decreases in HRV, increases in HR, and increases in both 7-day HRV CoV, 7-day HR CoV, and 7-day breath rate CoV were significantly associated with increased likelihood of a reported crash. On the between-person level, lower average HRV and higher HRV CoV were significantly associated with crash likelihood. Additionally, individuals who identify as female or non-binary had a higher likelihood of crashes. There was a statistically significant main effect of time, such that crash reports decreased slightly over the course of app usage. However, the odds ratio was effectively 1, indicating a negligible effect size. Finally, the prior-day crash fixed effect was a strong and significant predictor. Full MLM results for crashes are available in Table 3.

Next, we used MLM with random participant intercepts to model how daily morning biometric fluctuation predicted fatigue levels that evening (Fig. 2b). On the within-person level, decreases in HRV, increases in HR, and increases in 7-day HR CoV, HRV CoV, and breath rate CoV were associated with higher fatigue that evening. On the between-person level, higher average HR, HR CoV, and HRV CoV were significantly associated with elevated fatigue. Additionally, individuals who identify as female or non-binary reported higher average fatigue. There was a statistically significant main effect of time, such that fatigue reports decreased slightly over the course of app usage but, again. Finally, the prior-day fatigue fixed effect was a significant predictor. The full MLM model results for fatigue are available in Table 4.

**Table 1 | Dataset information per-user**

| Characteristic Per User | Median (IQR) |
|---|---|
| Number of Biometric Readings Provided | 125 (55, 257) |
| Length of Longest Consecutive Biometric Reporting Streak (days) | 20 (10, 44) |
| Number of Crash Episodes | 15 (5, 42) |
| Number of Fatigue Episodes (Dichotomized) | 76 (32, 171) |
| Number of Brain Fog Episodes (Dichotomized) | 27 (6, 83) |
| Duration of Crash Episodes (days) | 4 (2, 10) |
| Duration of Fatigue Episodes (days) | 18 (8, 49) |
| Duration of Brain Fog Episodes (days) | 6 (2, 19) |
| Days between Crash Episodes | 5 (3, 11) |
| Days between Fatigue Episodes | 1.68 (1.17, 2.78) |
| Days between Brain Fog Episodes | 3 (1, 6) |

Summary of User-Reported and Sensor-Derived Data Characteristics (per User; $n = 4244$). All values are presented as median (interquartile range). "Biometric readings" refer to the number of days on which each user contributed at least one biometric data point (e.g., heart rate, HRV, etc.). "Episodes" for fatigue and brain fog are defined using dichotomized variables used in predictive modeling. The longest streak is calculated based on consecutive days without missing biometric data. Days between episodes represent the average number of days between occurrences of a reported episode type.

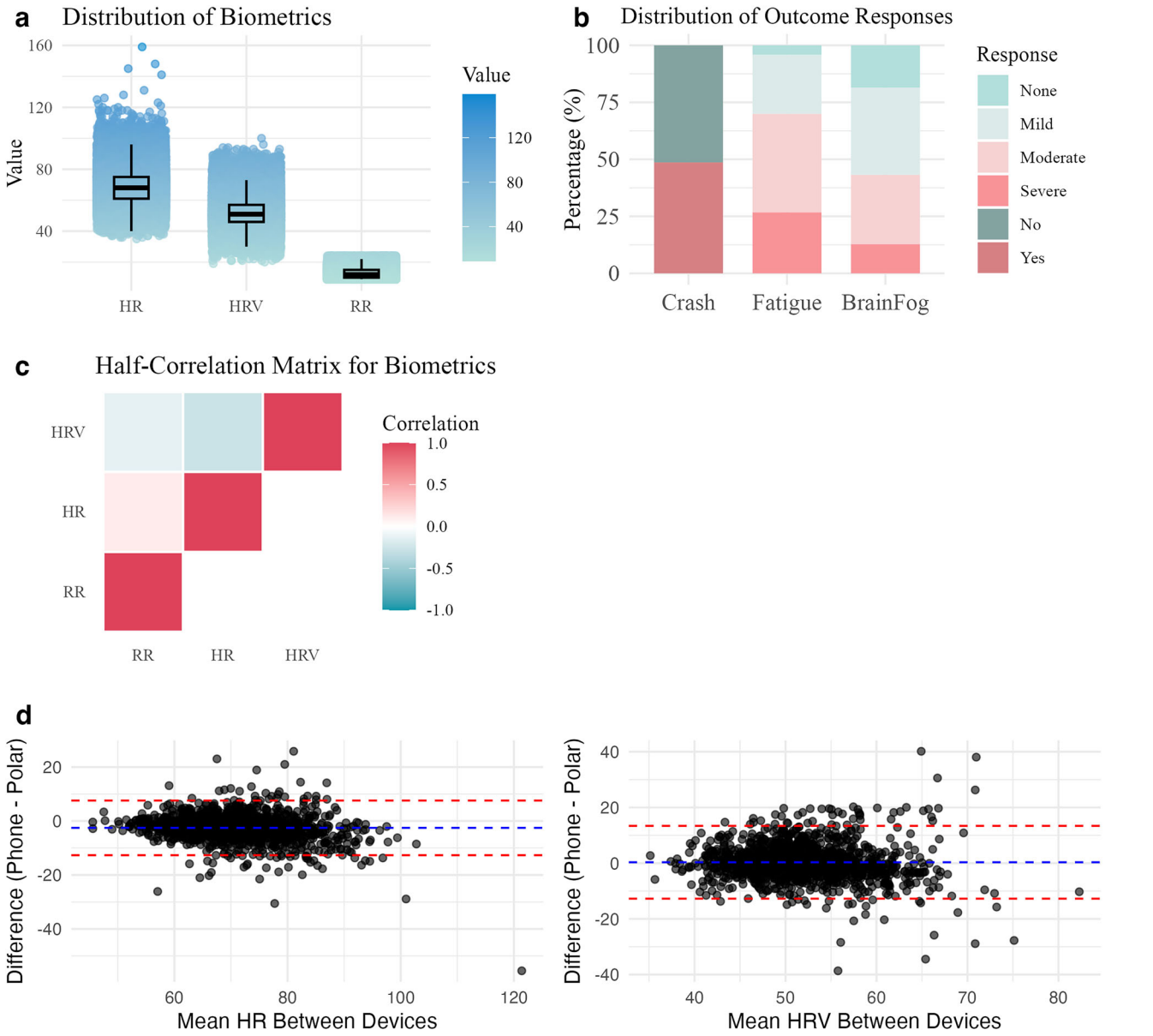


**Fig. 1 | Summary statistics and agreement between physiological measures.** **a** Scatterplot showing the distribution of heart rate (HR), heart rate variability (HRV), and respiration rate (RR) values collected across participants. Higher values are darker blue and lower values are lighter blue. **b** Stacked bar plot showing the distribution of response categories for three outcomes: crash (binary), and brain fog and fatigue (ordinal). Each bar represents the percentage of participants reporting distinct levels of symptom severity or presence. For the Brain Fog and Fatigue, response categories are displayed in teal (None), pale teal (Mild), pale rose (Moderate), and rose (Severe). For the binary outcome (crash), responses are shown in dark teal (No) and dark rose (Yes). **c** Correlation matrix heatmap among HR, HRV, and RR values. Color scale indicates direction and strength of Pearson correlations, with red indicating positive associations and blue indicating negative associations. **d** Bland–Altman–style plot comparing within-individual average HR estimates across devices for participants who contributed data from both the smartphone and Polar armband. Each point represents the difference between the device-specific mean HR values plotted against their average. The blue dashed line indicates the mean difference (bias), and red dashed lines show ±1.96 standard deviations.

Finally, we used MLM with random participant intercepts to model how daily morning biometric fluctuation predicted brain fog levels reported that evening (Fig. 2c). On the within-person level, decreases in HRV, and increases in HR and both 7-day HR CoV and 7-day HRV CoV were significantly associated with increased brain fog. On the between-person level, higher average HR and average breath rate, and higher HR CoV were significantly associated with increased brain fog. There was also a significant main effect of time, such that brain fog reports decreased across app usage. Individuals who identified as non-binary reported significantly higher average brain fog. The lag-1 brain fog fixed effect was also a strong and significant predictor. Full model results are presented in Table 5.

### Predictive model performance

Given the small, statistically significant effect sizes observed in the regression models, we implemented an exploratory walk-forward cross-validation (CV) approach to more thoroughly evaluate their predictive performance and practical utility. Specifically, we used a 5-fold walk-forward CV procedure to assess within-person temporal generalization and to compare the predictive performance of multilevel models across incremental specifications. In the walk-forward CV procedure, each participant's data was partitioned chronologically, such that training always preceded testing, allowing us to evaluate predictive accuracy on future observations from the same individuals. We focused on two primary model types that excluded

**Table 2 | Model statistics for predictors of crash, fatigue, and brain fog**

| MLM Model Comparison | | | | | |
|---|---|---|---|---|---|
| **Model** | **Marginal $R^2$** | **Conditional $R^2$** | **AIC** | **$\Delta\chi^2$** | ***p*** |
| **Crash** | | | | | |
| 1. Covariates | 0.003 | 0.571 | 259471.6 | NA | NA |
| 2. + Lag | 0.110 | 0.463 | 232435.6 | 27037.98 | <0.001 |
| 3. + Biometrics | 0.129 | 0.467 | 231871.7 | 587.93 | <0.001 |
| **Fatigue** | | | | | |
| 1. Covariates | 0.005 | 0.463 | 853929.5 | NA | NA |
| 2. + Lag | 0.127 | 0.395 | 808469.4 | 45462.08 | <0.001 |
| 3. + Biometrics | 0.155 | 0.404 | 807234.4 | 1259.07 | <0.001 |
| **Brain Fog** | | | | | |
| 1. Covariates | 0.003 | 0.594 | 727133.8 | NA | NA |
| 2. + Lag | 0.249 | 0.528 | 652419.6 | 74716.15 | <0.001 |
| 3. + Biometrics | 0.279 | 0.538 | 652041.1 | 402.50 | <0.001 |

Hierarchical multilevel model results showing marginal and conditional $R^2$ values (variance explained), AIC values (model fit), and $\chi^2$ statistics with p-values from likelihood ratio tests. Step 1 includes covariates only. Step 2 adds the prior-day symptom report (lag) variable. Step 3 adds biometric variables on top of covariates and lagged symptoms.

random effects in prediction (for subject-specific random intercepts and biometrics-only models see SI 6): (1) models including only lag-1 prior-day symptom reports, and (2) models that additionally incorporated morning biometric predictors. For this classification analysis, we used dichotomized fatigue and brain fog variables alongside the crash variable as outcomes (see SI Tables 4–5 for MLM results using dichotomized outcomes).

When only prior-day symptom predictors (i.e., lag-1 symptom reports) were included, models achieved AUC values were 0.78 for crash, 0.73 for fatigue, and 0.83 for brain fog. When morning biometric features were added to the prior-day symptom model, AUCs increased modestly across all outcomes, reaching 0.81 for crash, 0.74 for fatigue, and 0.85 for brain fog. These findings suggest that recent symptom history is a strong predictor of next-day symptom outcomes, with morning biometric data consistently offering incremental, and statistically significant (per DeLong's tests) improvements in model discrimination (See Table 6). Detection performance varied by outcome and generally did not differ between models with and without biometrics (see SI 4 for full metrics). For crash and brain fog, F1 scores were moderately higher for the prior-day symptom-only models and specificity was slightly higher for biometrics + prior-day symptom models. For fatigue, F1 score and specificity was equivalent for across both models. See SI Fig. 1 and SI Tables 9–12 for full classification metrics.

## Discussion

This study analyzed a large ($n = 4244$), longitudinal, high-frequency dataset of with LC, ME/CFS, or other energy-limiting chronic conditions. Participants provided 60-second morning biometric readings and self-reported evening symptom severity levels over time. Our findings revealed within-person level associations between morning biometric fluctuations and evening symptom reports, with model AUCs ranging from 73 to 85, depending on the inclusion of individual intercepts and previous-day symptom data. These results underscore the potential for improving care for individuals with complex chronic conditions through the targeted development of personalized, evidence-based remote physiological and behavioral monitoring systems.

Across all outcomes, within-person predictors of symptom changes were the strongest. We calculated daily fluctuations from baseline averages for morning heart rate variability (HRV), heart rate (HR), and respiration rate (RR), along with weekly changes in stability (CoV) for each biometric. RR was less predictive, contributing minimally to symptom prediction. Daily HRV and HR changes, along with 7-day biometric stability, emerged as key predictors, with higher HRV and lower HR associated with decreased risk of crashes, fatigue, and brain fog. These findings are in line with past research that has found that individuals with chronic illness, including LC and ME/CFS, show alterations in HRV and HR compared with normative populations[18,29–33]. More short-term variability in HRV and HR was linked to a higher probability of experiencing worsening symptoms, suggesting that fluctuations in cardiovascular dynamics over several days may destabilize daily symptom patterns. These results suggest that short-term fluctuations in HRV and HR dynamics, not just one-time metrics, are crucial for predicting symptom exacerbations. Furthermore, the intraclass correlation coefficients (ICCs) reported for each model ranged from 0.30 to 0.41, indicating a moderate degree of within-person variability. This level of variability is consistent with expectations for chronic conditions characterized by episodic or fluctuating symptomatology, and it supports the validity of modeling individual-level temporal changes in symptom severity.

To a lesser extent than within-person predictors, between-person baseline biometric patterns also demonstrated predictive value for symptom outcomes. Among these, the most consistent predictor was long-term variability in morning HR, where individuals with more stable morning HR patterns experienced fewer symptoms on average, including crashes, fatigue, and brain fog. This finding may reflect the role of HR stability as a marker of overall physiological resilience for people with complex chronic illnesses[34], where greater consistency in autonomic functioning reduces vulnerability to stressors that can trigger symptoms. Lower average morning HRV scores were linked to an increased likelihood of crashes but not fatigue or brain fog, suggesting that while HRV may be critical for predicting periods of acute stress (such as crashes), its influence on more chronic symptoms like fatigue and brain fog may be less pronounced. Conversely, higher average morning HR was associated with increased fatigue and brain fog, indicating that sustained elevations in HR could be a sign of prolonged physiological stress, contributing to chronic symptom experiences. Interestingly, individuals who identified as female or non-binary reported higher levels of fatigue, which could possibly reflect hormonal fluctuations related to the menstrual cycle[35–37].

In our exploratory predictive modeling analysis, we found that models combining morning biometric features from 60-second PPG assessments with prior-day symptom reports achieved AUC values of 0.74–0.85, which were significantly higher than those from models with prior-day symptom reports alone. This pattern underscores the strong predictive value of self-report data and the autocorrelative nature of symptom flare-ups, while also demonstrating that biometrics can provide incremental, statistically significant gains in predictive accuracy. However, the magnitude of improvement in overall AUC was modest. Prediction performance patterns varied across outcomes and were not uniformly enhanced by the addition of biometrics. F1 scores - the harmonic mean of recall and precision - for crash and brain fog were higher when using lag-only predictors, but sensitivity, which reflects the proportion of true symptom onsets correctly identified (and is inversely related to the false negative rate), was better when biometrics were included. For fatigue, detection scores did not significantly differ or benefit from the inclusion of biometrics. Thus, while the MLM models and AUC values in the predictive models showed superior performance when biometrics were included, the balance between correctly detecting onsets and avoiding false alarms, as reflected in the F1 score, did not improve with biometrics. This aligns with prior literature showing mixed results when using wearable-derived biometric data for symptom detection or health monitoring, where performance often varies depending on PPG recording variability, sensor quality, and individual differences (Roos and Slavich, 2023).

Although overall model discrimination was good, the models' ability to accurately detect symptomatic days was more limited. Detection performance was poor when models relied solely on biometric features without accounting for individual baselines, underscoring the need for personalized when using biometrics for real-world symptom prediction[38,39].

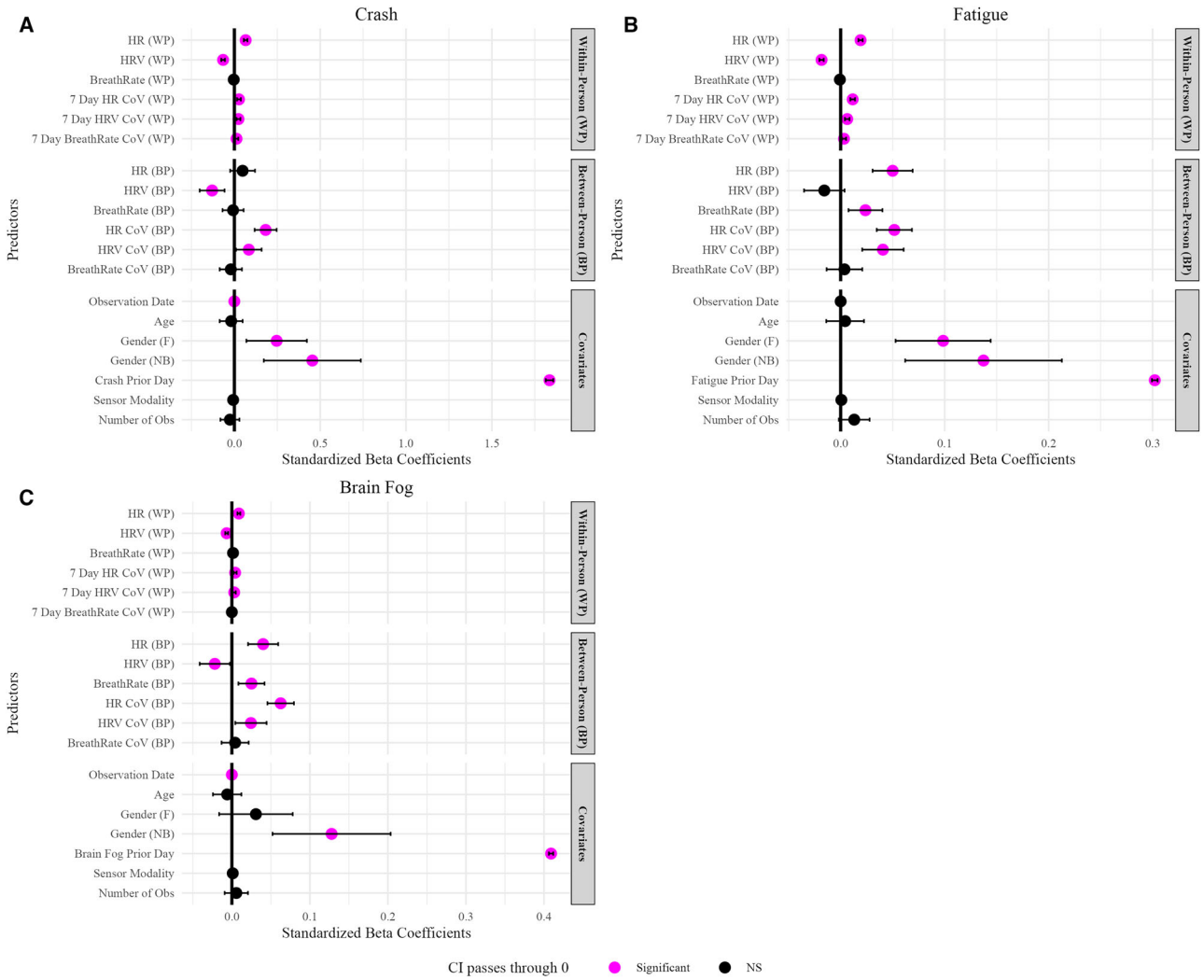


**Fig. 2 | Predictors of symptom severity across multilevel models.** Forest plots showing standardized effect estimates from multilevel regression models predicting three outcomes: **A** Standardized odds ratios for crash; **B** Standardized beta coefficients for fatigue severity; **C** Standardized beta coefficients for brain fog severity. Each point represents the effect size estimate with 95% confidence intervals. Predictors are categorized as: (1) Within-person (WP): person-mean centered values representing daily fluctuations around an individual's average (i.e., within-subject change); (2) Between-person (BP): grand-mean centered values representing each participant's overall average biometrics relative to the sample; (3) Covariates: demographic and contextual predictors (e.g., age, gender, device type). Variables include daily and 7-day coefficients of variation (CoV) for HR, HRV, and breath rate. Sensor Modality is coded as 1 for smartphone PPG sensor and 0 for Polar PPG. Pink circles indicate statistically significant predictors (CIs exclude 0); black circles indicate non-significance. Abbreviations: HR heart rate, HRV heart rate variability, WP within-person, BP between-person, CoV coefficient of variation, NB non-binary gender identity, F female.

Incorporating random effects improved detection performance across all outcomes, further supporting the use of individualized models that capture baseline symptom variation. False detection of symptomatic days was low for crash and fatigue, and relatively low for brain fog, suggesting models were generally specific, even when sensitivity was suboptimal. However, an important caveat in interpreting these findings is the presence of class imbalance, as non-symptomatic days were more prevalent in the dataset. This imbalance may have constrained model learning and impacted sensitivity in detecting symptom onset[40].

Mechanistic inferences are difficult given the uncontrolled nature of the dataset and the multisystem nature of illnesses such as LC and ME/CFS but exploring potential mechanisms behind the observed associations between low HRV, high HR, and symptom crashes remains important. HRV and HR are commonly recognized as proxy measures of autonomic nervous system function: increased SNS activity raises HR and reduces HRV, while PNS activity lowers HR and increases HRV[41]. One plausible mechanism involves the vagus nerve, which plays a key role in regulating inflammation and modulating central nervous system responses[42]. A meta-analysis found a consistent negative relationship between HRV and markers of inflammation[43], thought to be mediated through the cholinergic anti-inflammatory pathway[44].

Chronic inflammation is interrelated with the autonomic nervous system via this same cholinergic pathway and has been shown to decrease HRV and elevate resting HR[43,45,46]. In conditions such as LC and ME/CFS, persistent pathogens, reactivation of latent viruses, onset of autoimmunity, dysregulation of cortisol and other hormones, mitochondrial dysfunction and endothelial dysfunction have all been reported and can all lead to chronic pro-inflammatory responses with the potential to cause significant daily fluctuations in HR and HRV[47–51]. An important open question is the extent to which these autonomic signals can be disentangled from other concurrent physiological stressors, such as acute illness or infection, that may independently perturb HR and HRV. The potential for these mechanisms to not only cause chronic activation of inflammatory pathways, but also daily fluctuations in these activations could indeed explain how subsequent fluctuations in HRV and HR could be used to predict the emergence of crashes and other debilitating symptoms. Further research

**Table 3 | Autoregressive logistic MLM model predicting Crash**

| Crash | | | | |
|---|---|---|---|---|
| ***Predictors*** | ***Odds Ratios*** | ***std. Beta*** | ***Standardized CI*** | ***p*** |
| (Intercept) | 30.26 | 50.24 | 6.04–417.77 | **0.006** |
| HR (WP) | 1.01 | 1.07 | 1.06–1.08 | **<0.001** |
| HRV (WP) | 0.99 | 0.94 | 0.93–0.95 | **<0.001** |
| BreathRate (WP) | 1.00 | 1.00 | 0.99–1.01 | 0.468 |
| 7 Day HR CoV (WP) | 2.75 | 1.03 | 1.02–1.04 | **<0.001** |
| 7 Day HRV CoV (WP) | 2.07 | 1.03 | 1.02–1.04 | **<0.001** |
| 7 Day BreathRate CoV (WP) | 1.23 | 1.01 | 1.00–1.03 | **0.009** |
| HR (BP) | 1.00 | 1.04 | 0.97–1.11 | 0.286 |
| HRV (BP) | 0.98 | 0.87 | 0.81–0.94 | **<0.001** |
| BreathRate (BP) | 0.99 | 0.98 | 0.92–1.04 | 0.514 |
| HR CoV (BP) | 1135.64 | 1.20 | 1.13–1.27 | **<0.001** |
| HRV CoV (BP) | 11.12 | 1.09 | 1.02–1.17 | 0.013 |
| BreathRate CoV (BP) | 0.83 | 0.99 | 0.93–1.05 | 0.766 |
| Crash Prior Day | 6.43 | 2.24 | 2.21–2.26 | **<0.001** |
| Observation Date | 1.00 | 1.00 | 1.00–1.00 | **<0.001** |
| Number of Obs | 1.00 | 0.97 | 0.92–1.02 | 0.224 |
| Age | 1.00 | 0.98 | 0.92–1.05 | 0.584 |
| Gender (M) | 1.26 | 1.10 | 1.04–1.17 | **0.001** |
| Sensor Modality | 2.23 | 1.01 | 1.00–1.02 | 0.186 |
| **Random effects** | | | | |
| $\sigma^2$ | 3.29 | | | |
| $\tau_{00\ user\_id\_pk}$ | 2.09 | | | |
| ICC | 0.39 | | | |
| $N_{user\_id\_pk}$ | 3295 | | | |
| Observations | 325,656 | | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.129/0.467 | | | |

Final generalized multilevel model regression results for the crash outcome, reporting point estimates (odds ratios) alongside standardized effect sizes. WP = within-person; BP = between-person. Sensor modality is coded as 1 for smartphone PPG sensor and 0 for Polar PPG sensor.

**Table 4 | Autoregressive MLM model predicting Fatigue**

| Fatigue | | | | |
|---|---|---|---|---|
| ***Predictors*** | ***Estimates*** | ***std. Beta*** | ***Standardized CI*** | ***p*** |
| (Intercept) | 0.64 | 0.13 | −0.26–0.51 | **0.003** |
| HR (WP) | 0.00 | 0.02 | 0.02–0.02 | **<0.001** |
| HRV (WP) | −0.00 | −0.02 | −0.02 to -0.02 | **<0.001** |
| BreathRate (WP) | −0.00 | −0.00 | −0.00–0.00 | 0.491 |
| 7 Day HR CoV (WP) | 0.31 | 0.01 | 0.01–0.01 | **<0.001** |
| 7 Day HRV CoV (WP) | 0.12 | 0.01 | 0.00–0.01 | **<0.001** |
| 7 Day BreathRate CoV (WP) | 0.04 | 0.00 | 0.00–0.01 | 0.003 |
| HR (BP) | 0.01 | 0.05 | 0.03–0.07 | **<0.001** |
| HRV (BP) | −0.00 | −0.02 | −0.03–0.00 | 0.096 |
| BreathRate (BP) | 0.01 | 0.02 | 0.00–0.04 | 0.011 |
| HR CoV (BP) | 1.39 | 0.05 | 0.03–0.06 | **<0.001** |
| HRV CoV (BP) | 0.86 | 0.04 | 0.02–0.06 | **<0.001** |
| BreathRate CoV (BP) | 0.10 | 0.01 | −0.01–0.02 | 0.457 |
| Fatigue Prior Day | 0.31 | 0.31 | 0.31–0.31 | **<0.001** |
| Observation Date | −0.00 | −0.00 | −0.00–0.00 | 0.536 |
| Number of Obs | 0.00 | 0.01 | −0.00–0.03 | 0.140 |
| Age | 0.00 | 0.01 | −0.01–0.02 | 0.512 |
| Gender (M) | 0.07 | 0.04 | 0.02–0.05 | **<0.001** |
| Sensor Modality | 0.02 | 0.00 | −0.00–0.00 | 0.893 |
| **Random effects** | | | | |
| $\sigma^2$ | 0.33 | | | |
| $\tau_{00\ user\_id\_pk}$ | 0.14 | | | |
| ICC | 0.29 | | | |
| $N_{user\_id\_pk}$ | 3767 | | | |
| Observations | 459,609 | | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.155/0.404 | | | |

Final multilevel model regression results for the fatigue outcome, reporting point estimates alongside standardized effect sizes. Sensor modality is coded as 1 for smartphone PPG sensor and 0 for Polar PPG sensor.
*WP* within-person, *BP* between-person.

aimed at clarifying how these biological pathways interact over time may help refine the use of daily HRV and HR monitoring strategies to better manage symptom burden in conditions such as LC and ME/CFS.

This study has several notable limitations, which offer valuable insights for future research directions. Due to the retrospective study design, limited information was collected on demographics, making the generalizability of the findings uncertain. Future studies should aim to gather more detailed demographic information to assess the applicability of findings across different populations. Similarly, participants in this study reported that they met the WHO criteria for LC, but there were no standardized criteria for reporting ME/CFS. Additionally, this study included individuals with other energy-limiting conditions, which introduces variability and challenges in defining the sample population. This led us to refer to the dataset as representing individuals with complex chronic illnesses rather than a specific condition.

Participants measured biometrics using either smartphone cameras or armbands, which could introduce inconsistencies. Biometric indicators were derived from daily 60-s PPG assessments taken in the morning. While short heart-rate measurements via smartphone PPG recordings have been found reliable[52–54], future research using continuous day-long biometric data collection methods is likely needed to optimize the validity and reliability of physiological readings and thus potentially improve predictive model performance. Factors like time of day, temperature, skin tone, device type, software version, and recording length may have affected data accuracy. Participants were instructed to take measurements upon waking and while at rest, but the study's uncontrolled, real-world nature introduces potential variability. Although we assume compliance with these instructions, we cannot verify that all measurements were taken under true resting conditions, which may influence the interpretation of our findings. Along these lines, PPG-derived HR and HRV estimates may be less accurate than research-grade ECG or respiration belt measurements. Future work should further evaluate the validity and reliability of these biometric features compared to gold-standard physiological monitoring methods. Additionally, users who opted to use a dedicated wearable device rather than smartphone-based measurements may differ systematically in symptom burden or monitoring behavior, introducing potential selection bias. Although sensor modality was included as a covariate and analyses emphasized within-person changes, future studies should more directly

assess how device choice influences predictive performance and generalizability. Similarly, future work is needed to employ alternative sensor methods to better differentiate between fluctuations in specific symptoms. Incorporating additional signals, such as sleep patterns, temperature monitoring, and activity levels contextualized by individual baselines, may enhance the precision and specificity of symptom prediction. Despite these limitations, this study is the first to leverage data-driven, large-scale assessments of common symptoms among individuals with complex chronic illnesses.

Leveraging a natural intensive longitudinal data design from mobile health technologies, this study highlighted the potential of daily biometric monitoring to predict symptom fluctuations in individuals with complex chronic conditions. Within-person deviations in daily HRV and HR from a person's baseline and changes in their biometric weekly stability were robust predictors of crash, fatigue, and brain fog. While these models demonstrated promising predictive performance among existing users, further work is needed to enhance applicability to new populations. These findings suggest that digital health tools may support real-time symptom tracking and prediction, though further validation in prospective samples is needed to confirm generalizability.

**Table 5 | Autoregressive MLM model predicting Brain Fog**

| **Brain Fog** | | | | |
|---|---|---|---|---|
| ***Predictors*** | ***Estimates*** | ***std. Beta*** | ***Standardized CI*** | ***p*** |
| (Intercept) | 1.04 | 1.04 | 0.69–1.39 | **<0.001** |
| HR (WP) | 0.00 | 0.01 | 0.01–0.01 | **<0.001** |
| HRV (WP) | −0.00 | −0.01 | −0.01 to −0.00 | **<0.001** |
| BreathRate (WP) | 0.00 | 0.00 | −0.00–0.00 | 0.081 |
| 7 Day HR CoV (WP) | 0.11 | 0.00 | 0.00–0.01 | **<0.001** |
| 7 Day HRV CoV (WP) | 0.07 | 0.00 | 0.00–0.01 | **0.002** |
| 7 Day BreathRate CoV (WP) | 0.00 | 0.00 | −0.00–0.00 | 0.917 |
| HR (BP) | 0.00 | 0.04 | 0.02–0.06 | **<0.001** |
| HRV (BP) | −0.00 | −0.02 | −0.03–0.00 | 0.103 |
| BreathRate (BP) | 0.02 | 0.02 | 0.01–0.04 | **0.005** |
| HR CoV (BP) | 2.04 | 0.06 | 0.04–0.08 | **<0.001** |
| HRV CoV (BP) | 0.52 | 0.02 | 0.00–0.04 | 0.032 |
| BreathRate CoV (BP) | 0.09 | 0.00 | −0.01–0.02 | 0.555 |
| Brain Fog Prior Day | 0.41 | 0.41 | 0.41–0.42 | **<0.001** |
| Observation Date | −0.00 | −0.00 | −0.00 to −0.00 | **<0.001** |
| Number of Obs | 0.00 | 0.01 | −0.01–0.02 | 0.452 |
| Age | −0.00 | −0.00 | −0.02–0.01 | 0.609 |
| Gender (M) | 0.05 | 0.02 | 0.01–0.04 | **0.003** |
| Sensor Modality | 0.04 | 0.00 | −0.00–0.00 | 0.726 |
| **Random effects** | | | | |
| $\sigma^2$ | 0.29 | | | |
| $\tau_{00\ user\_id\_pk}$ | 0.16 | | | |
| ICC | 0.36 | | | |
| $N_{user\_id\_pk}$ | 3379 | | | |
| Observations | 402858 | | | |
| Marginal $R^2$ / Conditional $R^2$ | 0.279/0.538 | | | |

Final multilevel model regression results for the brain fog outcome, reporting point estimates alongside standardized effect sizes. Sensor modality is coded as 1 for smartphone PPG sensor and 0 for Polar PPG sensor.
WP within-person, BP between-person.

## Methods

No personal identifiable information was collected from this dataset. Thus, the dataset was determined to be exempt from human subject research by the Mount Sinai Program for the Protection of Human Subjects. However, all data presented here was derived from individuals who consented to provide their data for research purposes on the Visible application.

### Participants

People aged ≥18 years with self-identified complex chronic illnesses such as LC, ME/CFS, or people experiencing other causes of energy limitation who were using the Visible application and opted to share their data anonymously were eligible for inclusion. Participants ranged in age from 13 to 83 years (M = 46.3, SD = 12.4). The gender distribution was 80% female, 14% male, and 6% non-binary. Visible is a commercially available application that can be downloaded from major smartphone app stores, including both iOS and Android platforms. People self-identifying with LC were asked to confirm that they met the World Health Organization (WHO) definition of LC. Participants were included if they provided a minimum 14 biometric readings (see Fig. 3). Data used in this analysis was collected between August 2022 and April 2024.

### Biometrics

Resting heart rate (HR), heart rate variability (HRV), and respiration rate (RR) were measured using 60-s daily photoplethysmography (PPG) assessments collected in the morning. HRV follows a circadian rhythm, with higher parasympathetic tone and greater stability in the morning, making morning measurements particularly sensitive to physiologically meaningful deviations (Shaffer and Ginsberg, 2017).

In the unpaid version of the Visible application, users provided PPG readings using their smartphone camera. Camera-based PPG detects changes in blood volume in the skin by capturing light reflected from the fingertip placed over the camera lens. These measurements were processed by Spren, a commercial digital health company, which applies proprietary algorithms to derive HR, HRV, and RR from the camera-based optical signal. Raw PPG signals, intermediate image data, and signal-processing

**Table 6 | Mean ROC-AUC for crash, fatigue, and brain fog using walk-forward CV**

| **Mean ROC-AUC (Fixed Effects Only)** | | | |
|---|---|---|---|
| **Outcome** | **Prior Day Symptom Only** | **Biometrics + Prior Day Symptom** | **Model Comparison with DeLong's Significance Test ([test statistic range], p-value maximum)** |
| *Crash* | 0.78 | 0.81 | [12.9–14.6]; $p < 0.0001$ |
| *Fatigue* | 0.73 | 0.75 | [14.2–18.6]; $p < 0.0001$ |
| *Brain Fog* | 0.83 | 0.85 | [12.9–15.7]; $p < 0.0001$ |

ROC-AUC scores from models using fixed-effect predictors only (excluding random effects) for prior-day symptom–only models and biometrics + prior-day symptom models. DeLong's significance test was used to assess whether including biometrics resulted in a statistically significant improvement in model accuracy.

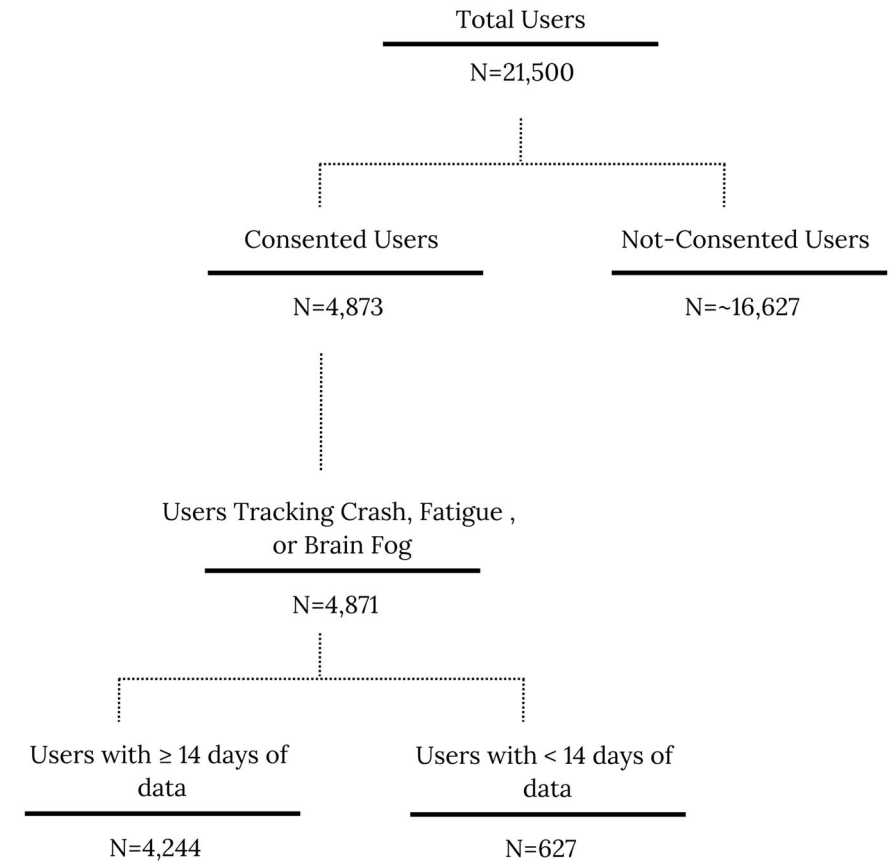


**Fig. 3 | CONSORT diagram.** A consort diagram detailing participant flow from total users to consented users, users tracking a primary outcome symptom, and users with the minimum amount of data required for inclusion.

parameters were not accessible to the study team. Participants using the subscription version of the Visible application had the option to measure HR and HRV using a Polar Verity Sense™ armband (Polar Electro, Inc., Kempele, Finland), a PPG-based optical heart rate sensor worn on the upper arm. The device was positioned with the sensor placed on the inner side of the armband to ensure firm skin contact. HR and HRV values from the Polar device were generated using Polar's proprietary optical processing algorithms, and raw PPG waveforms were not available for analysis. For both modalities, all preprocessing, including signal extraction, artifact handling, beat detection, and biometric computation, was performed internally by the respective platforms. At the study level, quality control was limited to excluding observations flagged as invalid by the device/application.

Users are instructed to complete the biometric recording in the morning, sitting or lying down and in the same position for each recording. HRV score was derived from calculating the root mean square of successive differences between heartbeats and scaling the score to receive an interpretable value between 0 and 100. Resting HR was calculated by measuring the number of heartbeats per minute. RR is calculated by detecting changes in heart rate related to breathing, where the heart rate increases during inhalation and decreases during exhalation. By leveraging these heart rate variations, the respiration rate is estimated. Additionally, the coefficient of variation (CoV) of an individual's biometric values over the previous seven days to assess physiological stability and variability.

To preserve the temporal integrity of the time series, we did not interpolate, impute, or otherwise stitch together non-consecutive days of biometric or symptom data. To ensure data quality, we applied adherence criteria for variables incorporating 7-day calculations, excluding rolling 7-day periods with more than two days of missing data. This approach preserved data integrity while maximizing the inclusion of usable data across participants.

### Symptom outcomes

Participants self-select which symptoms to track when creating an account with Visible and can enable additional symptoms at any time during app usage. Each evening, users receive a prompt to report their symptoms for that day (See Fig. 4).

A crash was defined within the app as periods of time when an individual's illness is significantly worse. The following explanation was provided to users: "Crashes usually occur as a part of post-exertional symptom exacerbation and affect your ability to carry out your usual activities. They normally last a few days. People use different words to describe these, such as flare-ups". Each evening, users are prompted to respond with either a "✓" or an "X" to the item labeled "Crash" under the question "What else happened today?" This prompt allows users to indicate whether they experienced a crash that day. Checkmarks were coded as 1 and Xs were coded as 0.

Each evening, users were prompted to log the severity of fatigue symptoms on a 0 to 3 scale (0 = no impairment; 1 = mild; 2 = moderate; 3 = severe). For the purposes of predictive classification, fatigue was dichotomized such that scores of 0 to 1 (no to mild symptoms) were coded as "0," and scores of 2 to 3 (moderate to severe symptoms) coded as "1." This dichotomization reflects the high prevalence of moderate fatigue symptoms in the sample, which necessitated grouping moderate responses with lower severity levels to ensure sufficient contrast in the outcome variable.

Each evening, users were prompted to log the severity of brain fog. Brain fog was self-reported by participants on a 0 to 3 scale (0 = no impairment; 1 = mild; 2 = moderate; 3 = severe). For predictive classification, brain fog was dichotomized with scores of 0 to 1 (no to mild symptoms) coded as "0," and scores of 2 to 3 (moderate to severe symptoms) coded as "1." This approach was chosen because moderate and severe brain fog symptoms were less common and grouping them helped balance the distribution across the dichotomous outcome. The difference in

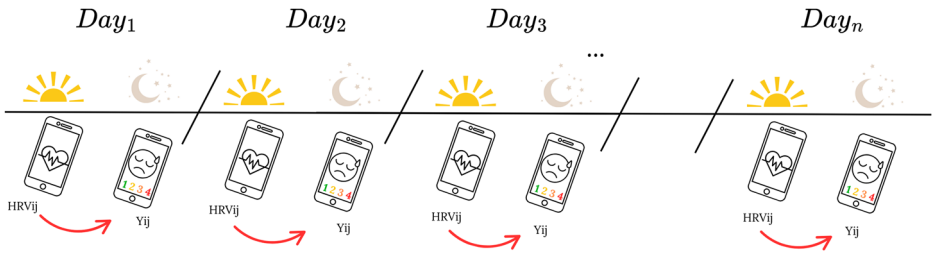


**Fig. 4 | Conceptual diagram of study design.** Adults with complex chronic illnesses completed a biometric assessment (heart rate, HR; heart rate variability, HRV; respiratory rate, RR) each morning, followed by a self-report symptom survey in the evening. The diagram illustrates the model structure, where morning biometric scores (e.g., HRV) predict same-day evening symptom severity. This approach allows for within-day analysis of the relationship between physiological measures and symptom changes.

dichotomization between fatigue and brain fog reflects the distinct underlying distributions of symptom severity in the dataset.

## Analysis

Analyses were performed in R version 4.3.0[55] using the lme4 package[56] for multilevel modeling and tidymodels[57] for data preprocessing for predictive modeling.

We conducted a generalized linear mixed-effects model analysis to predict same-day associations between morning biometrics and evening symptom reports (see equations 1 and 2). We disaggregated between-person (across participants) and within-person (within each participant) estimates of tonic and dynamic associations between biometrics and symptoms by partitioning predictor variances into between-person (grand-mean centered) and within-person (person-mean centered). Between-person predictors were time-invariant, while within-person predictors were time-varying. At the between-person level, we included all time-invariant predictors such as grand-mean biometric levels, stability scores, age, gender, and sensor modality.

To address the central question of whether biometrics provide incremental predictive value beyond prior-day symptom reports, we restructured our MLMs to follow a hierarchical, stepwise framework. Separate models were fit for each outcome, sequentially comparing: (1) Covariates-only models; (2) Models adding lag-1 symptom predictors; (3) Models further incorporating biometric features. Model comparisons were performed using likelihood ratio tests (ANOVA), with changes in model fit evaluated via marginal $R^2$, AIC, and $\chi^2$ statistics. Marginal $R^2$ reflects the proportion of variance explained by fixed effects alone, whereas conditional $R^2$ reflects the variance explained by both fixed and random effects, including between-person differences. This stepwise approach allowed us to directly quantify the added contribution of biometrics over and above both covariates and lagged symptom history.

On the within-person level, we included all time-varying predictors, including the group-mean centered morning biometric level and stability scores, observation date, and autoregressive lag-1 symptom value of the day prior. The lag-1 symptom variable is included in the models to control for serial correlation effects when an individual experiences multiple days of elevated or diminished symptomology.

Since we tested three primary outcomes (crash, fatigue, brain fog), we applied a Bonferroni correction to control the family-wise error rate. This yielded a corrected significance threshold of $p < 0.0167$. To remain even more conservative, we primarily interpret only those results meeting the more stringent $p < 0.01$ threshold.

Equation 1: Final Model (Binary generalized linear mixed effects model):

$$\left(\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right)$$

$$= \gamma_{00} + u_{0j} + \gamma_{10} * HR_{ij} + \gamma_{1j} * HR_{ij} + \gamma_{20} * HRV_{ij} + \gamma_{2j} * HRV_{ij} + \gamma_{30} * RR_{ij} + \gamma_{3j} * RR_{ij}$$

$$+\gamma_{40}*HR\_CoV_{ij}+\gamma_{4j}*HR\_CoV_{ij}+\gamma_{50}*HRV\_CoV_{ij}+\gamma_{5j}*HRV\_CoV_{ij} +\gamma_{90}*RR\_CoV_{ij}+\gamma_{6j}*RR\_CoV_{ij}$$

$$+\gamma_{7j}*date_{ij}+\gamma_{80}*age_{ij}+\gamma_{90}*gender_{ij}+\gamma_{10j}*Y_{i-1,j}+\gamma_{110}*source\_device_{ij}+\varepsilon_{ij}$$

Equation 2: Final Model (Continuous linear mixed effects model):

$$Y_{i,j}=\gamma_{00}+u_{0j}+\gamma_{10}*HR_{ij}+\gamma_{1j}*HR_{ij}+\gamma_{20}*HRV_{ij} +\gamma_{2j}*HRV_{ij}+\gamma_{30}*RR_{ij}+\gamma_{3j}*RR_{ij}$$

$$+\gamma_{40}*HR\_CoV_{ij}+\gamma_{4j}*HR\_CoV_{ij}+\gamma_{50}*HRV\_CoV_{ij} +\gamma_{5j}*HRV\_CoV_{ij}+\gamma_{90}*RR\_CoV_{ij}+\gamma_{6j}*RR\_CoV_{ij}$$

$$+\gamma_{7j}*date_{ij}+\gamma_{80}*age_{ij}+\gamma_{90}*gender_{ij}+\gamma_{10j}*Y_{i-1,j}+\gamma_{110}*source\_device_{ij}+\varepsilon_{ij}$$

In this multilevel model, $Y$ represents the outcome variable (e.g., crash, fatigue, brain fog) for observation $i$ within participant $j$. The group-level intercept, denoted as $\gamma_{00}$, represents the average outcome across all participants when all predictors are at their average values. The term $u_{0j}$ is the random intercept for participant $j$, capturing deviations in baseline outcomes for each participant compared to the overall average. This accounts for individual differences that are not explained by the fixed effects.

The model includes both within-person and between-person components. The within-person effects are captured by the linear terms $\gamma_{10}$ - $\gamma_{110}$. These terms represent the relationship between the fluctuations in each participant's biometrics and corresponding changes in symptom outcomes.

The between-person effects, represented by $\gamma_{1j}$- $\gamma_{10j}$, denote the relationships between each participant's overall (between-person) average levels of time-invariant variables and their average symptom outcomes. These terms capture how individual differences in long-term average biometric states relate to average symptom reporting.

In addition to the morning biometric variables, the model adjusts for other factors, including the observation date, age, gender, previous day's symptom reporting ($Y_{i-1,j}$), and the device used to collect biometric data. Finally, the residual error term $\varepsilon_{ij}$ accounts for unobserved variability in symptom outcomes at the individual level that is not explained by the model's predictors.

A potential concern is that, even with the inclusion of a lagged outcome term representing the prior day's symptom report to account for temporal dependencies, repeated occurrences of symptoms might bias our model. To address this, we conducted a sensitivity analysis that included only the first instance of a symptom report, coding subsequent consecutive reports as NA. The effects of the biometrics remained substantively unchanged. See SI Table 6–8.

## Predictive model evaluation

To evaluate within-person temporal generalization, we applied a stratified walk-forward 5-fold cross-validation (CV) procedure. Within each

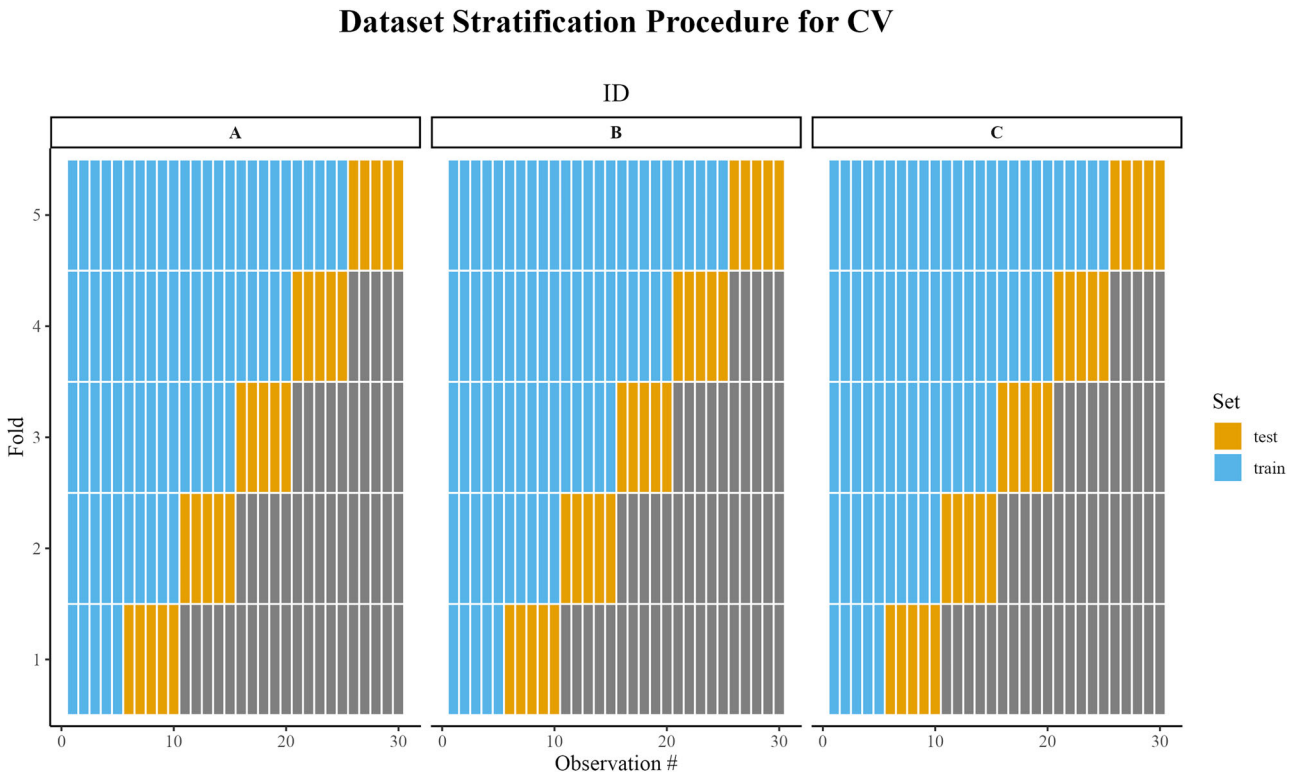


**Fig. 5 | Walk forward cross-validation procedure.** Figure illustrating the walk-forward stratified 5-fold cross-validation process for three example participants. The colored sections (blue and gold) represent training/testing for each fold within the larger training set (**A**-**C**).

participant, observations were ordered chronologically and divided into five approximately equally sized, sequential folds, ensuring that training data always preceded validation data in time. This walk-forward CV strategy was stratified by participant (Table 6), such that each participant contributed data to all training folds, enabling individualized modeling while preserving independence between training and validation observations. Additional details of the CV procedure are provided in Fig. 5. Visualization of missingness and biometric variability over time are provided in SI Figs. 2 and 3. On average, symptomatic days per user occurred at the following rates: 26% for crash reports, 64% for fatigue, and 44% for brain fog. To reduce class imbalance bias, we set a filter to exclude individuals who reported less than 5% of total logged days as a crash and those who logged more than 95% of days as fatigue. Classification was the primary objective, so the Fatigue and Brain Fog variables were dichotomized based on thresholds described above. Models were evaluated using ROC-AUC as the primary performance metric, alongside F1 Scores. ROC–AUC quantifies a model's ability to discriminate between symptom presence and absence, with higher values indicating better predictive performance and F1 scores summarize a model's ability to correctly identify symptom events by balancing precision.

Models were trained using a linear mixed-effects framework with lag-1 symptom variables alone or with morning biometric feature variables added. To evaluate the impact of individual-level variance on predictive performance, we generated predicted probabilities from the same fitted multilevel models in two ways: (1) excluding random effects, by setting re.form = NA, which yields population-level predictions based solely on fixed effects; and (2) including random effects, by specifying re.form = ~(1 | user_id), which incorporates subject-specific random intercepts to produce individualized predictions. Random-effects predictions are provided in SI 5 and are not shown in the main manuscript.

Finally, model performance within each fold was compared by generating 95% confidence intervals for AUC values using DeLong's method[58] with bootstrapping, implemented via the pROC package in R.

## Data availability

The data analyzed in this study were originally collected for use within a mobile application. To protect participant privacy, the data remain securely stored and are not publicly available. De-identified data may be made available upon reasonable request, provided that appropriate institutional and ethical approvals are obtained.

## Code availability

Data analysis was completed in R. The analysis code is available at upon reasonable request.

Received: 9 November 2024; Accepted: 3 March 2026;
Published online: 24 March 2026

## References

1. Thompson, E. J. et al. Long COVID burden and risk factors in 10 UK longitudinal studies and electronic health records. *Nat. Commun.* **13**, 3528 (2022).
2. Falk Hvidberg, M., Brinth, L. S., Olesen, A. V., Petersen, K. D. & Ehlers, L. The health-related quality of life for patients with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *PLoS ONE* **10**, e0132421 (2015).
3. Davis, H. E. et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine* **38**, 101019 (2021).
4. Humphreys, H., Kilby, L., Kudiersky, N. & Copeland, R. Long COVID and the role of physical activity: a qualitative study. *BMJ Open* **11**, e047632 (2021).
5. O'Brien, K. K. et al. Conceptualising the episodic nature of disability among adults living with Long COVID: a qualitative study. *BMJ Glob. Health* **8**, e011276 (2023).
6. Van Campen, C. L. M. & Visser, F. C. Higher resting heart rate myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) patients

compared healthy controls: relation stroke volumes. *Med. Res. Arch.* **10**, (2022).
7. Fang, S.-C., Wu, Y.-L. & Tsai, P.-S. Heart rate variability and risk of all-cause death and cardiovascular events in patients with cardiovascular disease: a meta-analysis of cohort studies. *Biol. Res. Nurs.* **22**, 45–56 (2020).
8. Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017).
9. Thayer, J. F., Ahs, F., Fredrikson, M., Sollers, J. J. 3rd & Wager, T. D. A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* **36**, 747–756 (2012).
10. Quer, G., Gouda, P., Galarnyk, M., Topol, E. J. & Steinhubl, S. R. Inter- and intraindividual variability in daily resting heart rate and its associations with age, sex, sleep, BMI, and time of year: Retrospective, longitudinal cohort study of 92,457 adults. *PLoS ONE* **15**, e0227709 (2020).
11. Mongin, D. et al. Decrease of heart rate variability during exercise: an index of cardiorespiratory fitness. *PLoS ONE* **17**, e0273981 (2022).
12. Cole, C. R., Blackstone, E. H., Pashkow, F. J., Snader, C. E. & Lauer, M. S. Heart-rate recovery immediately after exercise as a predictor of mortality. *N. Engl. J. Med.* **341**, 1351–1357 (1999).
13. Marques, K. C., Quaresma, J. A. S. & Falcão, L. F. M. Cardiovascular autonomic dysfunction in "Long COVID": pathophysiology, heart rate variability, and inflammatory markers. *Front. Cardiovasc. Med.* **10**, 1256512 (2023).
14. Bellocchi, C. et al. The interplay between autonomic nervous system and inflammation across systemic autoimmune diseases. *Int. J. Mol. Sci.* **23**, 2449 (2022).
15. Mooren, F. et al. Autonomic dysregulation in long-term patients suffering from Post-COVID-19 Syndrome assessed by heart rate variability. *Sci. Rep*. **13**, 15814 (2023).
16. da Silva, A. L. G. et al. Impact of long COVID on the heart rate variability at rest and during deep breathing maneuver. *Sci. Rep.* **13**, 22695 (2023).
17. Suh, H.-W., Kwon, C.-Y. & Lee, B. Long-term impact of COVID-19 on heart rate variability: a systematic review of observational studies. *Healthcare* **11**, 1095 (2023).
18. Barizien, N. et al. Clinical characterization of dysautonomia in long COVID-19 patients. *Sci. Rep.* **11**, 14042 (2021).
19. da Silva, R. B., Neves, V. R., Montarroyos, U. R., Silveira, M. S. & Sobral Filho, D. C. Heart rate variability as a predictor of mechanical ventilation weaning outcomes. *Heart Lung* **59**, 33–36 (2023).
20. Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M. & Suri, J. S. Heart rate variability: a review. *Med. Biol. Eng. Comput.* **44**, 1031–1051 (2006).
21. Nelson, B. W. et al. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *NPJ Digit Med.* **3**, 90 (2020).
22. Mather, J. D., Hayes, L. D., Mair, J. L. & Sculthorpe, N. F. Validity of resting heart rate derived from contact-based smartphone photoplethysmography compared with electrocardiography: a scoping review and checklist for optimal acquisition and reporting. *Front. Digit. Health* **6**, 1326511 (2024).
23. Chmiel, F. P. et al. Prediction of chronic obstructive pulmonary disease exacerbation events by using patient self-reported data in a digital health app: statistical evaluation and machine learning approach. *JMIR Med. Inform.* **10**, e26499 (2022).
24. Patel, M. L., Wakayama, L. N. & Bennett, G. G. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. *Obesity* **29**, 478–499 (2021).
25. Wood, J. et al. A smartphone application for reporting symptoms in adults with cystic fibrosis improves the detection of exacerbations: results of a randomised controlled trial. *J. Cyst. Fibros.* **19**, 271–276 (2020).
26. van Horck, M. et al. Early detection of pulmonary exacerbations in children with cystic fibrosis by electronic home monitoring of symptoms and lung function. *Sci. Rep.* **7**, 12350 (2017).
27. Hawks, Z. W. et al. Dynamic associations between glucose and ecological momentary cognition in Type 1 Diabetes. *NPJ Digit. Med.* **7**, 59 (2024).
28. Herrero, P. et al. Enhancing the capabilities of continuous glucose monitoring with a predictive app. *J. Diab. Sci. Technol.* **18**, 1014–1026 (2024).
29. Moshe, I. et al. Predicting symptoms of depression and anxiety using smartphone and wearable data. *Front. Psychiatry* **12**, 625247 (2021).
30. Kessler, R. C. et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol. Psychiatry* **21**, 1366–1371 (2016).
31. Proal, A. D. & VanElzakker, M. B. Long COVID or post-acute sequelae of COVID-19 (PASC): an overview of biological factors that may contribute to persistent symptoms. *Front. Microbiol.* **12**, 698169 (2021).
32. Proal, A. D. et al. SARS-CoV-2 reservoir in post-acute sequelae of COVID-19 (PASC). *Nat. Immunol.* **24**, 1616–1627 (2023).
33. Ryabkova, V. A., Rubinskiy, A. V., Marchenko, V. N., Trofimov, V. I. & Churilov, L. P. Similar patterns of dysautonomia in myalgic encephalomyelitis/chronic fatigue and post-COVID-19 syndromes. *Pathophysiology* **31**, 1–17 (2024).
34. Mensink, G. B. M. & Hoffmeister, H. The relationship between resting heart rate and all-cause, cardiovascular and cancer mortality. *Eur. Heart J.* **18**, 1404–1410 (1997).
35. Costeira, R. et al. Estrogen and COVID-19 symptoms: associations in women from the COVID Symptom Study. *PLoS ONE* **16**, e0257051 (2021).
36. Sund, M., Fonseca-Rodríguez, O., Josefsson, A., Welen, K. & Fors Connolly, A.-M. Association between pharmaceutical modulation of oestrogen in postmenopausal women in Sweden and death due to COVID-19: a cohort study. *BMJ Open* **12**, e053032 (2022).
37. Goodship, A. et al. Digital health app data reveals an effect of ovarian hormones on long COVID and myalgic encephalomyelitis symptoms. Preprint at https://doi.org/10.1101/2025.01.24.25321092 (2025).
38. Bent, B. et al. Engineering digital biomarkers of interstitial glucose from noninvasive smartwatches. *NPJ Digit. Med.* **4**, 89 (2021).
39. Hilton, C. B. et al. Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. *NPJ Digit. Med.* **3**, 51 (2020).
40. He, H. & Garcia, E. A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**, 1263–1284 (2009).
41. Berntson, G. G. et al. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology* **34**, 623–648 (1997).
42. Tracey, K. J. The inflammatory reflex. *Nature* **420**, 853–859 (2002).
43. Williams, D. P. et al. Heart rate variability and inflammation: a meta-analysis of human studies. *Brain Behav. Immun.* **80**, 219–226 (2019).
44. Tracy, L. M. et al. Meta-analytic evidence for decreased heart rate variability in chronic pain implicating parasympathetic nervous system dysregulation. *Pain* **157**, 7–29 (2016).
45. Ackland, G. L. et al. Autonomic regulation of systemic inflammation in humans: a multi-center, blinded observational cohort study. *Brain Behav. Immun.* **67**, 47–53 (2018).
46. Haensel, A., Mills, P. J., Nelesen, R. A., Ziegler, M. G. & Dimsdale, J. E. The relationship between heart rate variability and inflammatory markers in cardiovascular diseases. *Psychoneuroendocrinology* **33**, 1305–1312 (2008).
47. Peluso, M. J. & Deeks, S. G. Mechanisms of long COVID and the path toward therapeutics. *Cell* **187**, 5500–5529 (2024).

48. Iwasaki, A. & Putrino, D. Why we need a deeper understanding of the pathophysiology of long COVID. *Lancet Infect. Dis.* **23**, 393–395 (2023).
49. Klein, J. et al. Distinguishing features of long COVID identified through immune profiling. *Nature* **623**, 139–148 (2023).
50. Turner, S. et al. Long COVID: pathophysiological factors and abnormalities of coagulation. *Trends Endocrinol. Metab.* **34**, 321–344 (2023).
51. Appelman, B. et al. Muscle abnormalities worsen after post-exertional malaise in long COVID. *Nat. Commun.* **15**, 1–15 (2024).
52. van Dijk, W., Huizink, A. C., Oosterman, M., Lemmers-Jansen, I. L. J. & de Vente, W. Validation of photoplethysmography using a mobile phone application for the assessment of heart rate variability in the context of heart rate variability-biofeedback. *Psychosom. Med.* **85**, 568–576 (2023).
53. Bánhalmi, A. et al. Analysis of a pulse rate variability measurement using a smartphone camera. *J. Healthc. Eng.* **2018**, 4038034 (2018).
54. Peng, R.-C., Zhou, X.-L., Lin, W.-H. & Zhang, Y.-T. Extraction of heart rate variability from smartphone photoplethysmograms. *Comput. Math. Methods Med.* **2015**, 516826 (2015).
55. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, 2013).
56. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Soft.* **67**, 1–48 (2015).
57. Kuhn, M. & Wickham, H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. https://www.tidymodels.org (2020).
58. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).


## Acknowledgements

We gratefully acknowledge the dedicated Visible user community for their generous participation in this research. We also thank the editor and reviewers for their thorough review and constructive critique of the manuscript. The authors did not receive specific funding for this study. A.A. and D.P. received research support from the Steven & Alexandra Cohen Foundation.

## Author contributions

A.A. conducted the analyses and wrote the main manuscript text. A.S. contributed substantially to writing the main manuscript. A.I., H.M., J.T.M., and A.P. contributed to the theoretical conceptualization and editing of the manuscript prior to publication. R.P. and P.C. provided input on the methods and analytical approach. H.L. reviewed and edited the manuscript prior to submission. M.O. contributed to the analytic framework and its formulation. D.P. led the study design and manuscript conceptualization.

## Competing interests

A.A. received consulting fees from Visible Health Inc. not related to the production of this manuscript. R.P., P.C., and H.L. are current or former employees of Visible Health Inc. and did not play a role in the study design or decision to publish the manuscript. A.S., H.M., J.T.M., A.P., M.O., and D.P. declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41746-026-02543-3.

**Correspondence** and requests for materials should be addressed to Annie Aitken or David Putrino.