5-20-2021

Predicting circadian phase across populations: a comparison of mathematical models and wearable devices

Yitong Huang

Caleb Mayer

Philip Cheng
Henry Ford Health, PCHENG1@hfhs.org

Alankrita Siddula

Helen J. Burgess

See next page for additional authors

Recommended Citation
Huang Y, Mayer C, Cheng P, Siddula A, Burgess HJ, Drake C, Goldstein C, Walch O, and Forger DB. Predicting circadian phase across populations: a comparison of mathematical models and wearable devices. Sleep 2021.

SLEEPJ, 2021, 1–11

doi: 10.1093/sleep/zsab126
Advance Access Publication Date: 20 May 2021
Original Article

ORIGINAL ARTICLE

Predicting circadian phase across populations: a comparison of mathematical models and wearable devices

Yitong Huang[1], Caleb Mayer[2,], Philip Cheng[3,], Alankrita Siddula[4], Helen J Burgess[5,], Christopher Drake[3,], Cathy Goldstein[6], Olivia Walch[6] and Daniel B. Forger[2,7,8,*]

[1]Department of Mathematics, Dartmouth College, Hanover, NH, USA, [2]Department of Mathematics, University of Michigan, Ann Arbor, MI, USA, [3]Henry Ford Health System, Detroit, MI, USA, [4]Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL, USA, [5]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA, [6]Department of Neurology, University of Michigan, Ann Arbor, MI, USA, [7]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA and [8]Michigan Institute for Data Science, University of Michigan, Ann Arbor, MI, USA.

*Corresponding author. Daniel B. Forger, Department of Mathematics, Department of Computational Medicine and Bioinformatics, Michigan Institute for Data Science, University of Michigan, 2074 East Hall, 525 East University, Ann Arbor, MI 48104. Email: forger@umich.edu.

Abstract

From smart work scheduling to optimal drug timing, there is enormous potential in translating circadian rhythms research results for precision medicine in the real world. However, the pursuit of such effort requires the ability to accurately estimate circadian phase outside of the laboratory. One approach is to predict circadian phase noninvasively using light and activity measurements and mathematical models of the human circadian clock. Most mathematical models take light as an input and predict the effect of light on the human circadian system. However, consumer-grade wearables that are already owned by millions of individuals record activity instead of light, which prompts an evaluation of the accuracy of predicting circadian phase using motion alone. Here, we evaluate the ability of four different models of the human circadian clock to estimate circadian phase from data acquired by wrist-worn wearable devices. Multiple datasets across populations with varying degrees of circadian disruption were used for generalizability. Though the models we test yield similar predictions, analysis of data from 27 shift workers with high levels of circadian disruption shows that activity, which is recorded in almost every wearable device, is better at predicting circadian phase than measured light levels from wrist-worn devices when processed by mathematical models. In those living under normal living conditions, circadian phase can typically be predicted to within 1 h, even with data from a widely available commercial device (the Apple Watch). These results show that circadian phase can be predicted using existing data passively collected by millions of individuals with comparable accuracy to much more invasive and expensive methods.

Statement of Significance

Previous work has shown that light measurements from a research-grade wearable device (e.g. Actiwatch) coupled with mathematical models provide an inexpensive and noninvasive approach to predict human circadian phase. However, ubiquitous consumer-grade wrist-worn wearable devices (e.g. Apple Watch or Fitbit) do not report light measurements, but rather activity. Here we examined estimating circadian phase using activity data as an input, an approach that provides phase predictions with comparable accuracy to using light exposure for people living under normal conditions, but outperforms the predictions using light from Actiwatch in a shift worker population. We compared circadian predictions from multiple mathematical models to ground truth dim light melatonin onset (DLMO), across normal and shift worker samples, to identify the optimal model. This sets the stage for deploying widely available commercial devices to predict human circadian phase on a global scale.

Key words: circadian rhythms; mathematical models; actigraphy; wearable data

Submitted: 20 December, 2019; **Revised:** 22 March, 2021

Introduction

Circadian clocks govern the timing of physiological processes in many organisms [1–5]. The central human circadian pacemaker is known to influence academic/work performance, alertness, and various health conditions, including diabetes, cancer, and mental diseases [6–9]. Knowing a person's circadian timing or phase can lead to targeted therapeutic interventions and better health treatments. Much research has focused on the mechanism of circadian rhythms on a molecular scale or in carefully controlled laboratory conditions [1, 5, 10, 11]. Therefore, a long-term goal of the field of chronobiology has been to translate this deep mechanistic understanding of circadian timekeeping to predicting human circadian phase in real life.

Mathematical modeling of the human circadian central pacemaker provides a passive and noninvasive alternative to predict circadian phase. As light is the dominant input to entrain the human circadian clock, it features prominently in efforts to model and predict circadian phase [12–18]. Kronauer et al. introduced a van der Pol limit cycle model with a nonlinearity of degree 7 (higher-order model) in 1999, which takes light measurements as the direct input and predicts core body temperature [16]. Based on the same model structure, Forger et al. proposed a simplified model with similar accuracy to the higher-order model [19]. Though light is the primary stimulus to the human circadian pacemaker, the circadian system is also affected by nonphotic stimuli, such as the sleep–wake cycle, activity, and associated behaviors [20, 21]. Therefore, a revised model was proposed in 2007, where a nonphotic component of the activity-rest cycle was added to the higher-order model [22]. Hannay et al. recently developed a separate model formalism for the human circadian pacemaker based on the network of neurons that control circadian pacemaking [23].

Recent work has shown that mathematical models of the human circadian clock coupled with the passive recordings of light levels from wearable devices such as research-grade actigraphy can be used for circadian assessment outside of controlled laboratory environments [24, 25]. Woelders et al. showed that the standard deviation of DLMO (dim light melatonin onset) predictions is 1.14 h under normal living conditions [25]. Stone et al. further tested the ability of models to predict circadian phase in workers on a rotating shift, finding a mean absolute error of 0.95 h between observed and predicted aMT6s acrophase with the nonphotic model and 1.19 h with the higher-order model during a night shift rotation [24]. Though the nonphotic model has been validated in a field setting, the performance of other models still remains unknown. In particular, the Hannay model has not been validated in an ambulatory setting.

Moreover, many widely available wearable devices already owned by millions of individuals do not record light measurements but record activity data. Activity, a circadian-regulated physiological outcome, has been used as a measure for the circadian rhythm in field studies [25–27]. Ambulatory activity is largely dependent on an individual's sleep–wake timing, and the sleep–wake timing is closely related to the circadian phase [28]. The objectives of this study were therefore to evaluate the accuracy of circadian prediction models with and without the inclusion of light information. In particular, we wanted to assess the quality of these predictions at two extremes: (1) in individuals subject to circadian disruption and (2) in data sets without direct measurements of light. In the first case, we performed this assessment in a published data set with night shift workers on nonrotating schedules, whose highly variable sleep–wake schedules make predicting circadian phase extremely difficult. To address the second case, we next applied circadian prediction models to activity data recorded by the multisensory wrist-worn consumer device (Apple Watch) and compared the circadian phase estimates against gold standard DLMO measurements.

In this study, we show that: (1) activity can be an indicator for light; (2) using activity data processed in a variety of ways as inputs to the circadian models can predict circadian phase with similar accuracy to light measurements for subjects who live under normal conditions; (3) model predictions from activity data significantly outperform those from light for a sample of nonrotating shift workers; (4) there is no significant difference between the performance of four circadian models, when the same input is used; and (5) in a pool of 20 nonshift worker subjects wearing the widely available Apple Watch, circadian phase can be predicted within 1 h using activity data alone.

Methods

Data sets

Our analyses include three data sets: a sample of day workers with both research-grade actigraphy (Actiwatch-L and Actiwatch Spectrum, Philips Respironics, Inc. Bend, OR) data and in-lab DLMO [29], a published data set of night shift workers with both research-grade actigraphy data and in-lab DLMO data [30, 31], and an unpublished data set from a nonshift working population with a multisensory wrist-worn consumer device (Apple Watch, Series 2 and 3) and in-lab DLMO data. To better demonstrate the results across data sets, the threshold to define the ground-truth DLMO was defined as the time when the melatonin concentration exceeded the mean plus two standard deviations of three low consecutive daytime salivary melatonin values for all data sets. In the following, we will describe the data sets.

Day workers

Thirteen participants from Chicago included in this data set wore an Actiwatch-L on their nondominant wrist and the Actiwatch Spectrum around the neck for a week. The data collected from the wrist were used in our analysis. Three subjects who had data collected for less than 5 days were excluded here, leaving us with a data set consisting of 10 subjects. Among these 10 subjects, 7 subjects participated in the study in summer, and the others participated in winter. No data was missing among these 10 subjects, and the median light level was 12.57 lux. The subjects reported to the laboratory on the 8th day, and DLMO measurements were collected via salivary melatonin measurements every 30 min from 7 h before to 3 h after their approximate bedtime. All subjects were full-time office workers, who were instructed to maintain their usual sleep and work schedule [29]. Activity (in activity counts as the cumulative sum of motion measured by a triaxial accelerometer) and light measurements (in lux) were collected in 30-s intervals for each of the subjects. This dataset has been described in further detail in a previous publication [29].

Night shift workers

The shift worker data set consisted of 27 night shift workers, who lived in Michigan and worked at least 3 night shifts per week during the study. Fifteen of 27 shift workers were recruited in summer, and the others participated in winter. All 27 participants were instructed to wear an Actiwatch Spectrum that measured light levels (lux) and activity counts (30 s or 1 min epochs) for 7–14 days. Actiware software (Philips Actiware 6.0.9) sleep/wake classifications were used to simulate the nonphotic model [22]. The median light level across subjects was 12.88 lux. The average percentage of missing light data was 8.80%; however, no obvious missing data (i.e. missing consecutively longer than 15 min) was found. Following ambulatory monitoring (mean = 12.41 days, *SD* = 4.35 days), participants reported to the laboratory after finishing a night shift, and 24-h DLMO measurements were collected in a controlled laboratory setting. Thirteen subjects kept wearing the device during the DLMO collection process, whereas the other 14 subjects' data ended when they reported to the laboratory. Further details of the DLMO collection process were described in a previous publication [30]. As shift workers have a highly variable DLMO across the 24-h day, this data set provided a wide spread of DLMO times with a standard deviation of 4.02 h.

Apple watch data set

For this data set, 20 healthy nonshift workers were instructed to wear an Apple Watch for 7–14 days before coming into the University of Michigan Sleep and Circadian Research Laboratory for DLMO assessment. Six subjects were recruited in summer, and the others participated in spring or fall. Due to the limitations of the device, only activity data were collected and light data were not obtained in this study. We should note that the unit of activity data from the Apple Watch is different from that of actigraphy, though both actigraphy and the Apple Watch measure activity using the triaxial accelerometer. Moreover, due to the battery life of the Apple Watch, approximately 6–8 h of data were lost every 1–2 days. Since the devices were usually removed (i.e. charged) during sleep, the subjects were assumed to be inactive (i.e. activity data were assumed to be 0) during the period of missing data. Salivary collection began 6 h prior to the participants' habitual bedtime; samples were collected with salivettes every 30-minutes until their bedtime (13 samples total).

Description of models

We examined the performance of four models: two versions of light-based mathematical models (the higher-order model of degree 7 proposed by Kronauer et al. in 1999 and the simpler model of degree 3 by Forger et al. in 1999) [16, 19]; the nonphotic model that accounted for an additional nonphotic term to the light-based higher-order model [22]; and Hannay's physiology-based circadian neural network model [23]. All four models predict the core body temperature rhythm, and estimated DLMOs can be obtained from the published relationship between CBTmin (minimum of core body temperature) and DLMO (DLMO = CBTmin – 7) [32, 33]. Further details of the models are described in the Supplement.

Model implementation

For each subject, at least 7 days of light and activity data (or activity data only from the Apple Watch) were recorded until the day of the DLMO assessments. Due to the difference of devices, the data were reported in different intervals, and we therefore resampled the data at a one-minute rate. To model the circadian response to ambulatory light data, both the light-based models and Hannay's physiology-based model use light measurements as inputs, whereas the nonphotic model requires not only the light measurements but also the activity-rest patterns provided by the sleep–wake indicator from the Actiwatch. We then replaced light measurements by activity data (or activity-derived light in the Supplement) to explore if activity can be used to predict the circadian phase.

Determining initial conditions is essential when implementing ordinary differential equation models. Due to the lack of prior information on subjects' circadian rhythms, we assume that every subject lives with a circadian rhythm regulated by 16 h of lightness (800 lux) and 8 h of darkness (0 lux) before entering the study. Both the light-based and the nonphotic models contain two state variables x and x_c exhibiting a limit cycle in the phase plane, where x reflects the endogenous core body temperature, and x_c is required to achieve the limit cycle mathematically. Simulating the light-based models or the nonphotic model gives us 24 pairs of x and x_c, where one pair of x and $x_c$was chosen every hour. Since Hannay's model describes the mechanism of the coupled oscillators in the SCN, it contains two different state variables, R and ψ, which represent the collective amplitude and the average phase of the oscillators, respectively. Therefore, the simulation of Hannay's model provides 24 pairs of the collective amplitude R and the collective phase ψ, and each pair represents the circadian state at every hour. The initial conditions of each subject were then chosen from these 24 pairs of representative circadian states, based on the hour of the timestamp of the first data point.

All codes used to implement the models and perform the analysis are available at https://github.com/ojwalch/predicting_dlmo.

Cosinor analysis of activity data

For the day workers and night shift workers data sets, activity acrophase was obtained by fitting a single harmonic cosine wave with a period of 24 h to the available activity data. The cosinor analysis was performed in MATLAB (Mathworks, R2018a), where nonlinear regression models were fitted with the *fitnlm* function.

Sleep timing as a proxy of DLMO

Sleep timing has previously been used as a proxy of circadian phase in day work settings [28, 34, 35]. For subjects in the day workers data set and the Apple Watch data set who live in a regular routine, we used habitual bedtime during the recording period as a proxy for DLMO, where the bedtime was determined by the mean bedtime estimated from actigraphy data during the days of data collection. For the night shift workers' data set, we separated the sleep timing following night shifts (average bedtime of the daytime sleep estimated from actigraphy) and the sleep timing on nonworkdays (average bedtime of the nights without night shifts) as possible indicators for DLMO. Based on past studies [28, 36, 37], estimated DLMO can be obtained by subtracting 2 h from the sleep timing. In addition, the role of

sleep–waking timing played in the models was further examined in the Supplement.

Scaling factor for Apple watch data

Since the proprietary activity data (i.e. Apple steps) is in a different unit than activity counts from actigraphy, a scaling factor is needed to scale the Apple Watch activity data to compensate for the device difference before feeding the data into the models. Therefore, the scaled activity was used to simulate the models for the Apple Watch data set. The scaling factor (=30/mean(steps)) was found by optimizing the mean absolute error between measured DLMO and estimated DLMO simulated from the higher-order model for five randomly selected subjects. This factor was then applied to the unseen testing data that contains 15 subjects.

Statistical analysis of model comparison

To assess if the model outputs from light are consistently different from those from activity, we used the sign test on the absolute errors obtained from different models with activity input and light input, where the prebuilt sign test was available from *signtest* in MATLAB. To compare the performance of the four models given the same input, we used the Friedman test from the function *friedman* in MATLAB.

Results

Relationship between light and activity

Binning light levels and activity measurements recorded from actigraphy in 10 participants from the day worker dataset revealed a general trend of increasing light levels for increasing activity counts (Figure 1A). However, once 500 lux is reached, a plateau is evident (i.e. different activity levels correspond with approximately the same light levels). To show the robustness of this feature across data sets, Figure 1B shows a similar relationship in a shift worker dataset of 27 subjects, in which a plateau appears once 500 lux is reached. Since full-time office workers tend to maintain a regular daily routine, the interquartile range (represented by the shaded area) is smaller for the day worker dataset. Fortunately, the dynamic stimulus processor built into the model of the circadian pacemaker is particularly sensitive to lower light levels (<50 lux), which suggests this plateau might not be as important as originally proposed [15, 18, 38, 39]. Figure 1A and 1B show that activity is correlated to light, and moreover, we are able to estimate light levels from activity counts using a simple piecewise linear function as well as more complex approach machine Learning (ML) techniques, for which more details can be found in the Supplement. Hence, in addition to light measurements, we explored the use of activity data to predict circadian phase.

Day worker dataset

We first examined the performance of circadian models using a regular, nonshift worker data set that contains 10 subjects who have a small range of DLMOs (range: 19:43–22:16, mean = 20:54, *SD* = 0.85 h). It is intuitive to fit a 24-h periodic signal to activity for day workers, since they follow a similar routine every day. Here we fitted a cosine wave to 7 days of data. A difference of 4.47 $\pm$ 2.28 h was observed between activity acrophase and DLMO, in line with the previous work showing that cosinor analysis gives an average difference of 4.6 h between activity acrophase and DLMO in a sample of subjects living under normal conditions [25]. However, a wide range of activity acrophases was observed (*SD* = 3.41 h), despite the fact that subjects had a narrow range of DLMO values (*SD* = 0.85 h). Moreover, adding 4.47 to the activity acrophase still yields predicted DLMOs with a mean absolute error of 3.05 h and a standard deviation of 2.45 h. Therefore, simply fitting a sinusoidal function to activity levels does not appropriately reflect circadian phase.

We next used all three van der pol limit cycle models and Hannay's physiological model to simulate the circadian response to ambulatory light data. Figure 2 and Table 1 demonstrate that activity data as an input to the model provides reasonable predictions when compared to light input: with activity, 20% more subjects reach an error within 60 min and 10% more subjects reach an error within 120 min from the higher-order model. Though the simpler model provides similar results, 10% and 20%

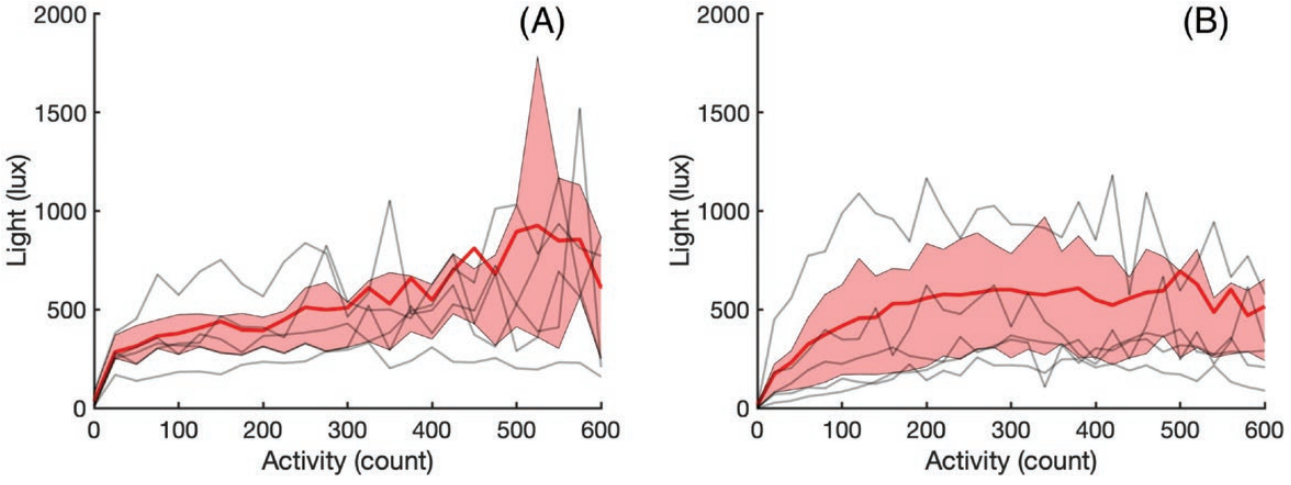

Figure 1. Relationship between light and activity. The relationship between light and activity is found from (A) the day worker data set and (B) the shift worker data set. The mean values are plotted in red, and the interquartile range is shaded. The gray lines show the data from 5 randomly selected individuals from each data set.

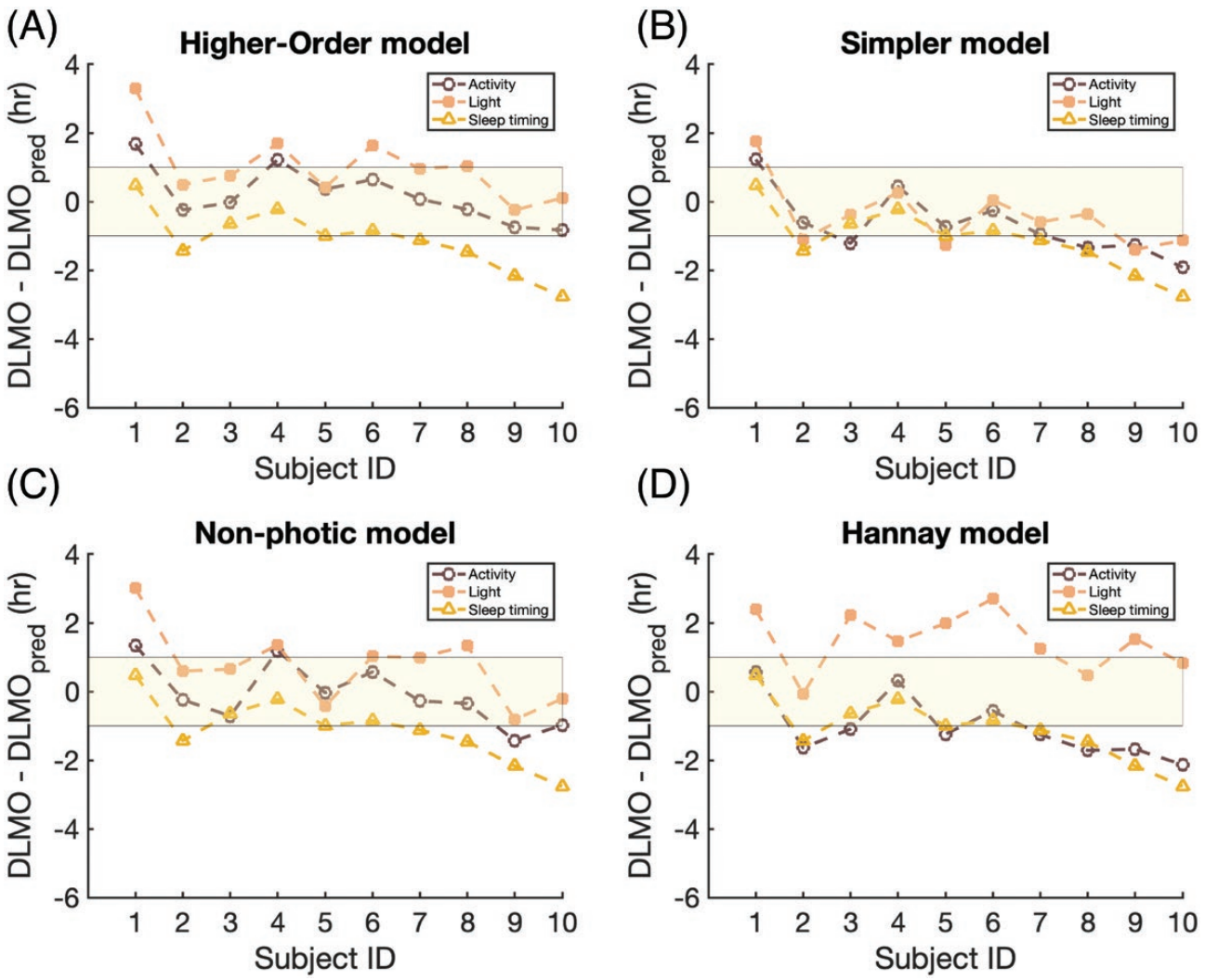

Figure 2. Day worker data set: the error between actual DLMO and predicted DLMO ($DLMO_{pred}$) for (A) higher-order model, (B) simpler model, (C) nonphotic model, and (D) Hannay model. Circles and squares represent predicted DLMOs simulated using four different models (higher-order model, simpler model, nonphotic model and Hannay model) with 2 different inputs (light and activity) respectively. Predicted DLMO from sleep timing (calculated by subtracting 2 h from the habitual bedtime) is marked in triangles. Prediction error of 1 h is shaded.

Table 1. Day worker data set: summary of prediction error in hours for higher-order, simpler, nonphotic and Hannay models with different inputs. * denotes the best model prediction for each model input based on the mean absolute error

Device	*n*	Method	Mean absolute error (h)	SD of absolute error (h)	Prediction within 60 min	Prediction within 120 min
Input: activity						
Actiwatch	10	Higher-Order Model	0.60534*	0.5321	80%	100%
Actiwatch	10	Simpler Model	0.99667	0.49524	50%	100%
Actiwatch	10	Nonphotic Model	0.713	0.4984	70%	100%
Actiwatch	10	Hannay Model	1.2198	0.59212	30%	90%
Input: light						
Actiwatch	10	Higher-Order Model	1.0627	0.95116	60%	90%
Actiwatch	10	Simpler Model	0.82967*	0.57496	50%	100%
Actiwatch	10	Nonphotic Model	1.041	0.7846	60%	90%
Actiwatch	10	Hannay Model	1.4938	0.85675	30%	70%

more subjects have an absolute error under 120 min using activity via the nonphotic model and Hannay's model, respectively.

To further compare the results from using activity and light as inputs, we use the sign test on the absolute errors obtained from different models with activity input and light input, yielding a nonsignificant difference between outputs from activity and light (p-value > 0.05). Despite the fact that the Hannay model yields the lowest number of subjects predicted within

1 h, an analysis of the Friedman test shows that all models perform similarly when the same input is used (p-value > 0.05).

Examining their daily schedule reveals that the subjects' habitual bedtime ranged from 22:25 and 00:40. A proxy of sleep timing calculated by subtracting 2 h from the habitual bedtime was then compared to DLMO, which yields a prediction with a mean absolute error of 0.97 h with a standard deviation of 0.64 h (Figure 2). The Friedman test also showed that no statistically significant difference was found between the performance of the proxy of sleep timing and that of the light-based and physiological models.

Shift worker dataset

We then validate the models on a shift worker population whose highly variable sleep–wake schedule poses challenges to circadian phase estimation. Though cosinor analysis shows that the 24-hr component is significant (p-value < 0.001), a difference of 6.02 ± 2.49 h was found between activity acrophase and DLMO, which is not consistent with the difference between activity acrophase and DLMO observed in the day workers dataset above.

In general, larger mean absolute errors are obtained from model predictions in the shift worker population. Table 2 shows that at least 20% more subjects can be predicted within 2 h when activity is used instead of light. Figure 3 shows that activity input provides comparatively more accurate estimates no matter which model is applied, as the DLMO timings of more individuals are predicted with an error under 3 h. Figure 4 further compares the estimated DLMO with the measured DLMO. Bias correction from Lin's concordance, ranging from 0 to 1, measures how deviated the best fit line is from the 45 degree line through the origin, where a bias correction closer to 1 means the best fit line is closer to the diagonal line. Figure 4 shows that all models with activity input perform better than those with light input, as a larger bias correction was obtained for every model with activity input. We then apply the sign test on absolute errors derived from models with activity inputs and light inputs, yielding a statistically significant difference between activity-input predictions and light-input predictions (p-value < 0.0001). Moreover, the Friedman test shows that four models behave similarly when activity is the input (p-value > 0.5); and no statistically significant difference between models was found when light is used as input as well (p-value >0.5).

To show the contribution of these mathematical models of the human circadian clock, we evaluated the ability of sleep timing to serve as a proxy of DLMO in this population. As the subjects arrived at the lab for DLMO measurements after a night shift, we first compared the habitual sleep timing following night shifts to the measured DLMO. A mean absolute error of 6.54 h was found between sleep timing-derived DLMO and true DLMO, and only 3 out of 27 subjects were predicted within 3 h. Bias correction from Lin's concordance was 0.228, which further indicates that habitual sleep timing following night shifts is a poor indicator for circadian phase. We then compared the average sleep timing on nonworkdays (i.e. the nights without night shifts) against DLMO (Figure 5). Though average sleep timing on nonworkdays is more likely to reflect the circadian phase than the sleep timing following night shifts, it still results in a mean absolute error of 3.86 h, where 29.6% of the subjects can be predicted within 2 h and less than half of the subjects can be predicted with an error below 3 h. Moreover, the linear relationship between sleep timing on nonworkdays and measured DLMO is not statistically significant (p-value > 0.05). These results suggest that mathematical models of the human circadian clock have exceptional and unique contributions to predicting circadian phase, especially in the shift worker population.

Of particular interest, the collective amplitude of oscillators (where the period of the oscillators is ~24 h) is related to the magnitude of circadian disruptions; specifically, more circadian disruptions exist as the amplitude decreases. Here, we found the minimum of collective amplitude simulated throughout available data. Figure 6 shows a statistically significant positive linear relationship between error and amplitude (p-value < 0.001), which implies the circadian clock slows as the amplitude decreases. We should note that the two subjects with the smallest amplitude either had relatively shorter data (approximately 7 days of data) or remained continuously active (activity count exceeds 300) until 2 h before the reported DLMO. The positive linear relationship remained statistically significant (p-value < 0.05), even when these two subjects were excluded. Thus, we find that it is more challenging to predict the circadian phase of shift workers with more circadian disruption.

Apple watch dataset

Thus far, the assessments of circadian models used datasets in which activity and light data were collected from research-grade actigraphy. Having shown that activity alone can be used to predict circadian phase, we applied the same modeling approach to data collected from a device currently owned and

Table 2. Shift worker data set: summary of prediction error in hours for higher-order, simpler, nonphotic and Hannay models with different inputs. * denotes the best model prediction for each model input based on the mean absolute error

Device	*n*	Method	Mean absolute error (h)	SD of absolute error (hr)	Prediction within 60 min	Prediction within 120 min
Input: activity						
Actiwatch	27	Higher-Order Model	2.6969	2.3127	18.5%	51.9%
Actiwatch	27	Simpler Model	2.7873	2.3257	18.5%	55.6%
Actiwatch	27	Nonphotic Model	2.5055*	2.1566	22.2%	51.9%
Actiwatch	27	Hannay Model	2.7625	2.262	25.9%	44.4%
Input: light						
Actiwatch	27	Higher-Order Model	3.7232	2.4123	14.8%	25.9%
Actiwatch	27	Simpler Model	3.5955*	2.5514	22.2%	25.9%
Actiwatch	27	Nonphotic Model	3.7072	2.1198	7.4%	25.9%
Actiwatch	27	Hannay Model	3.8115	2.4399	22.2%	22.2%

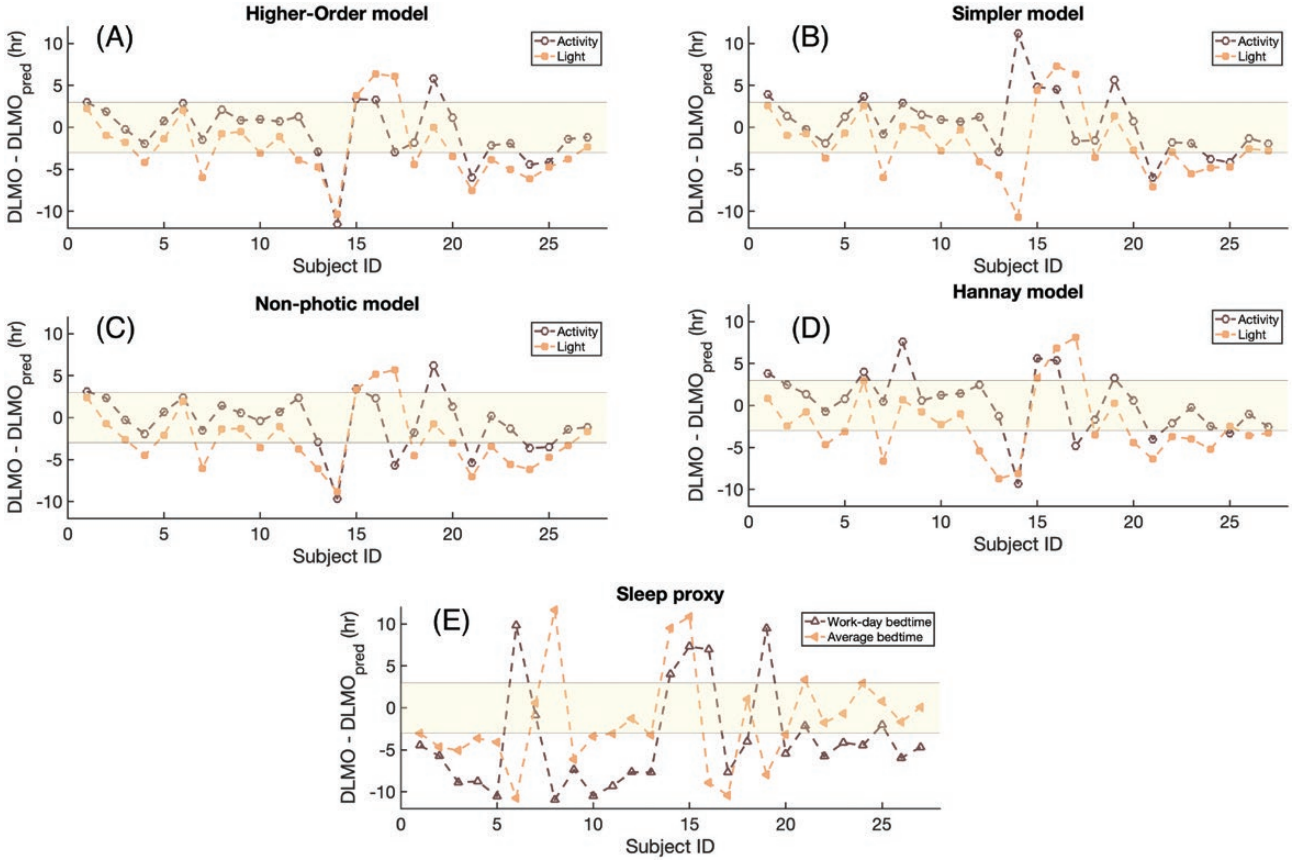

Figure 3. Shift worker data set: the error between actual DLMO and predicted DLMO ($DLMO_{pred}$) for (A) higher-order model, (B) simpler model, (C) nonphotic model, (D) Hannay model and (E) sleep timing. (A-D) Predicted DLMOs are simulated using four different models (higher-order model, simpler model, nonphotic model and Hannay model) with 2 different inputs (light and activity). (E) Predicted DLMOs from sleep timing are calculated by subtracting 2 h from two different bedtimes (sleep timing following a night shift and the average sleep timing during the ambulatory recording period). Prediction error of 3 h is shaded.

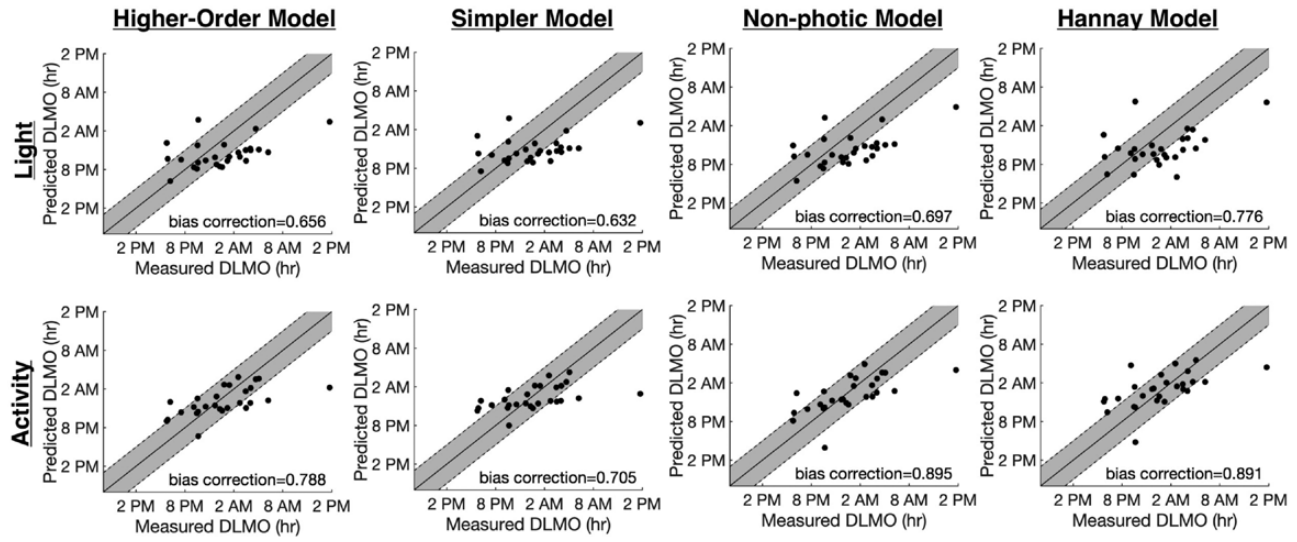

Figure 4. Shift worker data set: measured DLMO vs. predicted DLMO. Predicted DLMOs are simulated using four different models (higher-order model, simpler model, nonphotic model and Hannay model) with 2 different inputs (light and activity). Bias correction of Lin's Concordance measures how deviated the best fit line is from the 45 degree line through the origin, where a bias correction closer to 1 means the best fit line is closer to the diagonal line. Prediction error of 3 h is shaded.

worn on a daily basis by millions of consumers. Here, we tested the higher-order model with data collected from one commercial device, the Apple Watch, which was worn by 20 subjects for one week before laboratory DLMO measurement. After applying a scaling factor (trained from five randomly selected subjects) to the activity data, scaled activity was then inputted into the model. This approach gives a mean absolute error of 0.809 h with a standard deviation of 0.736 h for the 5 training subjects

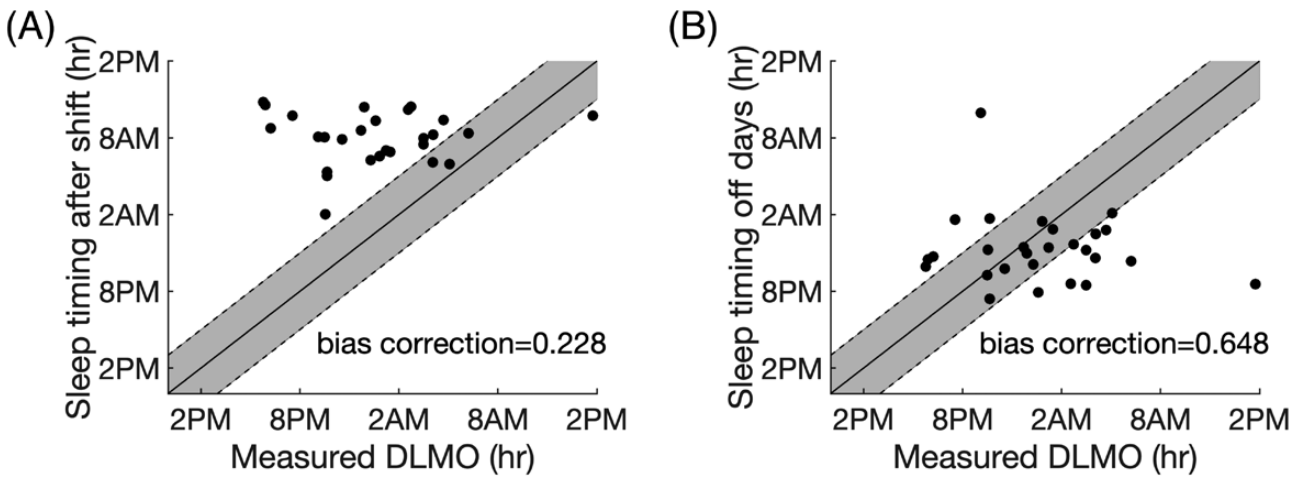

Figure 5. Shift worker data set: sleep timing as a proxy of DLMO. (A) Habitual bedtime following night shifts is used as a proxy of DLMO. (B) Average bedtime during the nights without night shifts is used as a proxy of DLMO. Bias correction of Lin's Concordance is reported and prediction error of 3 h is shaded.

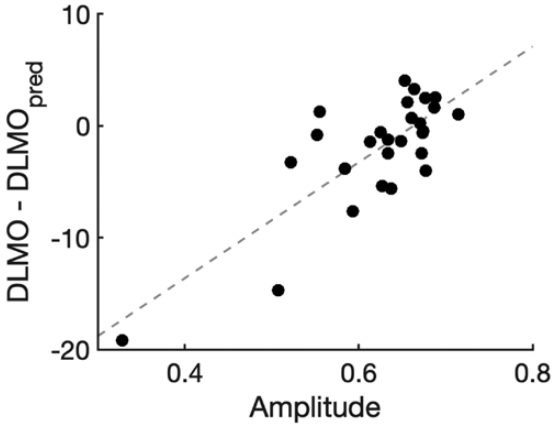

Figure 6. Role of amplitude in the shift worker data set. For each subject in the shift worker data set, we found the minimum of the collective amplitude of the oscillator population from Hannay model. A significant linear relationship between error and amplitude was found (p-value < 0.001), and this linear relationship remained significant even when two subjects who had the lowest amplitude were excluded (p-value < 0.05).

and a mean absolute error of 0.964 h with a standard deviation of 0.724 h for the unseen testing data (15 subjects), in line with previous work showing that circadian phase can be predicted approximately with an error of 1 h for individuals living in a regular life setting [25]. Figure 7 shows that 11 of those 15 testing subjects can be predicted with an error under 1 h, and the two largest errors were obtained by two subjects who had particularly early DLMOs (at 1745 and 1817). These data show that the circadian phase could be accurately measured even with a commercial device, even one that needs to be offline for approximately 8 h a day while charging or sleeping.

Subjects collected from Apple Watch are nonshift workers and have habitual bedtime from 22:00 to 24:10. Subtracting 2 h from habitual sleep timing provides predictions with a mean absolute error of 1.02 h and standard deviation of 0.84 h, though bias correction indicates that the performance from sleep timing is slightly worse than that from using the circadian model (Figure 7). Together with previous results, we show that sleep timing can be used as a proxy of DLMO for the regular population whose habitual bedtime is between 22:00 and midnight, but sleep timing after night shifts, in particular, is not sufficient to predict DLMO.

Discussion

There is great interest in determining circadian phase in real world situations. Previous studies have shown that this is possible. For example, Wu et al. used circadian biomarkers from the skin to accurately predict phase within 3 h from a single sample [40]; Wittenbrink et al. shows that circadian biomarkers in blood are able to provide phase estimates of 40 healthy young participants with a median absolute error under 1 h, where 12 individuals were conducted in a carefully controlled constant routine protocol and another group of 28 subjects was measured in a home environment [41]. Here, we present an exhaustive assessment of a noninvasive approach for predicting circadian phase in the real world. Our analysis shows that activity data from actigraphy may be as useful as raw light data in predicting circadian phase, at least when collected by a wrist-worn watch.

Many past analyses of circadian rhythms in actigraphic data have used cosinor analysis to fit activity patterns [25, 42]. In cosinor modeling, activity is taken as an output of the circadian clock from which circadian time can be estimated. In contrast to cosinor analysis, limit cycle oscillator models model the mechanics of the system, taking a zeitgeber history as the input and yielding a phase prediction found by cumulatively summing the phase shift achieved by the input at every point in time provided.

All of the limit cycle oscillator models we have compared in this paper take a lighting history as their primary input, with no designated input for activity (with the exception of activity-derived sleep–wake patterns in the nonphotic model). In passing activity into the model as a proxy for light, we are allowing activity to phase shift the model of the clock as light does. The observed performance, which is comparable or better than light itself, suggests several interpretations. One is that activity is possible to be better able to describe the actual ocular light levels experienced by the participants than wrist-worn light sensors. Another is that activity as an input to the model can capture both the phase shifting effects of light and an additional phase shifting effect of exercise on the circadian clock [43].

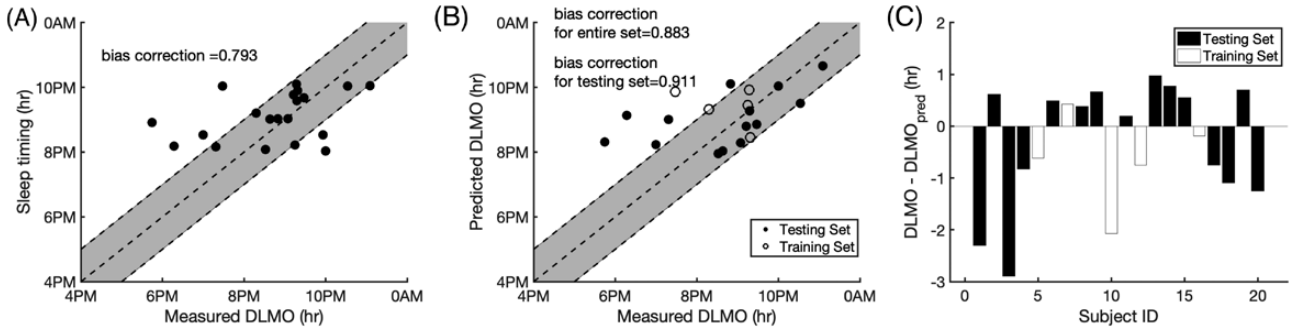

Figure 7. Apple Watch data set: testing the model on a widely available wearable device. (A) Sleep timing acts as a proxy of DLMO, which provides a mean absolute error of 1.02 h for 20 subjects. (B, C) The model predictions are simulated using scaled activity data (derived by applying a scaling factor to the activity data (i.e. step counts)) from Apple Watch as an input (more details in text), where the scaling factor was trained using 5 randomly selected subjects. This approach provides a mean absolute error of 0.809 h between measured DLMO and predicted DLMO for the five training subjects, and a mean absolute error of 0.964 h for 15 unseen testing subjects (filled in black). Prediction error of 1 h is shaded.

This intriguing possibility suggests the need for future work and further collection of high-quality lighting and activity data. In addition, while self-selected light exposure and exercise are both driving inputs to the clock, they are also outputs of the clock state. Post-hoc analyses of the observed sleep–wake and activity patterns could be used to tune and personalize parameters in the model yielding better overall predictions.

One of the models we tested has previously validated a sleep–wake input that is separate from the effect of light. When we used this sleep–wake input alone, it was not able to predict circadian phase as well as the methods we tried. We also used a synthetically generated activity profile to account for the effects of sleep and wake, and to determine if that alone could be used as an input to the model to predict circadian phase. That, as well, was not able to yield as accurate predictions as the methods we propose (See Supplemental Information and Figures S2–S4). This indicates that the activity levels recorded by the device add increased predictive ability to the models. Using these activity levels directly is also the most straightforward method possible to predict circadian phase in field settings. This approach opens circadian studies up to the millions of activity recordings currently being generated by wearables.

We conclude that the circadian phase can be generally estimated to 1 h in a normal life setting, which agrees with the accuracy obtained from the plasma biomarkers of DLMO [41]. As expected, model predictions contained larger errors in individuals with circadian disruption (i.e. night shift workers) than nonshift workers. Moreover, the analysis of 20 Apple Watch users further points to the potential use of a wider range of wearable technology in clinical populations with a more critical need for accurate and timely assessments of circadian phase.

Further work is also needed to improve the accuracy of mathematical models. Our results showed that for less than 30% of the night-shift subjects was the DLMO prediction within 1 h of the actual value. Lower circadian amplitude obtained from night shift workers may increase the vulnerability of their circadian system to perturbations, which might be the main reason for the difficulty in predicting circadian phase. The existing mathematical models do not account for the relationship between amplitude and period (as shown in Figure 6), which may be considered in future work with the aim of constructing personalized models to improve predictions. Moreover, we can observe that most subjects from the Apple Watch dataset can be predicted with high accuracy; much of the error was dominated by two subjects who had particularly early DLMOs (at 17:45 and 18:17). It is possible that their early DLMOs could be due to a short circadian period, as is found in Advanced Sleep Phase Syndrome. If so, improved model predictions could be found by using a shorter circadian period in the model. In addition, the discrepancy in accuracy between day workers and night shift workers may also result from the lack of prior information on the initial circadian states for the night shift workers, since this prior information can be reasonably estimated from the habitual sleep/wake schedule of typical day workers. Less information regarding initial conditions can be inferred from the more disrupted sleep schedule of shift workers, and in this case, more days of data are needed to obtain accurate and reliable estimates of circadian phase in this population.

Our work provides evidence that one week of data is sufficient to provide a relatively accurate estimate for individuals without a shift work schedule. However, the length of data required to reach similar accuracy for a highly variable population needs to be further tested. Moreover, the relationship between CBTmin and DLMO found in carefully controlled human studies has not been validated in the shift worker population. Incorporating clinical data or other subject characteristics might improve model predictions in shift workers, which should also be considered in the future.

Our study is subject to some limitations that should be noted. As mentioned above, it is possible that activity was able to match the performance of light as an input because it better approximated actual light exposure. The possible errors in light measurements using wearable devices, such as the differences between ocular and wrist measurements, and the effects of nonphotic cues such as food intake are not accounted for in our modeling. Wrist-worn light measurements are known to underreport absolute light measurements when compared to calibrated, laboratory standard photometers [44, 45]. In addition, even though subjects were instructed not to cover the watch, apparel choices (e.g. sleeves covering the devices) could have biased the light recording. These are important factors to consider, and all could have contributed to the enhanced value of activity relative to light data in predicting circadian phase.

The fact that activity measurements may vary by season or geographical location may also affect activity in ways that are not accounted for in the current model structure. Future work could address this by applying the models to a larger dataset

and potentially rescaling the activity data input to account for these group differences. In addition, each dataset included in our analysis consisted of less than 30 subjects, so future work deployed at a broader scale is needed. One particular weakness is the scaling factor applied to the Apple Watch users was tested on 15 subjects, and the robustness of this factor should be tested further in a larger scale of dataset.

We also should note that the Apple Watch loses approximately 6–8 h of data every day, when subjects typically remove their devices to charge during night. Here, we considered subjects as inactive during the data loss period in our study. As data loss occurred primarily at night, this choice could upweight the relative importance of sleep and wake timing in the model predictions. However, the effects of the timing and length of the data loss on circadian predictions need to be further investigated, since the sensitivity of the circadian pacemaker differs throughout the day [46]. To extend this work to populations, further calibrated comparisons of activity measurements from various wearable devices must take place.

We have shown that activity could serve as an alternative input to models which predict circadian phase when light information is not available. These results suggest the potential of using widely available consumer-grade wearable devices to track the circadian rhythm in a longitudinal and real-life setting. This does not mean, however, that researchers studying circadian rhythms should stop collecting light information. Light remains the primary circadian time cue to synchronize the human circadian clock. Yet that activity performs so well in place of light is noteworthy. Indeed, it may be that using activity as an input to the clock model captures the independent phase shifting effects of both light and exercise as zeitgebers. Our result is of particular interest given a surprising recent finding that supplementing wrist-level light recordings with eye-level recordings does not necessarily improve the accuracy of predictions based on the wrist light recordings alone [47]. We encourage researchers to seek to collect the highest quality light and activity data in order to fully understand the effects of these two inputs on the circadian clock.

In sum, our results indicate that activity measurements can produce reasonably accurate estimates of DLMO in individuals under normal living conditions, with relatively worse but still useful results for night shift workers. These results suggest new avenues for utilizing noninvasive wearable data for predicting circadian phase in both clinical and real-life settings.

Acknowledgments

This work was supported by National Science Foundation (NSF) Grant 1714094, National Institutes of Health (NIH) Grant 1R43CA236557-01A1, and National Institutes of Health (NIH) Grant K23 HL138166.

Disclosure Statement

Financial disclosure: This work was supported by NSF Grant 17140459, NIH Grant 1R43CA236557-01A1, and NIH Grant K23 HL138166. O.W. is the CEO and D.B.F. the CSO of Arcascope, a company that makes software for circadian rhythms. The University of Michigan is a part owner of Arcascope. Arcascope did not sponsor this research.

Non-financial disclosure: none.

References

1. Czeisler CA, *et al*. Sleep and circadian rhythms in humans. *Cold Spring Harb Symp Quant Biol*. 2007;**72**:579–597.
2. Moore-Ede M, *et al*. *The Clocks That Time Us: Physiology of the Circadian Timing System*. Cambridge, MA: Harvard University Press; 1982.
3. Nagoshi E, *et al*. Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell*. 2004;**119**(5):693–705.
4. Refinetti R, *et al*. The circadian rhythm of body temperature. *Physiol Behav*. 1992;**51**(3):613–637.
5. Sassone-Corsi P. Molecular clocks: mastering time by gene regulation. *Nature*. 1998;**392**(6679):871–874.
6. Gale JE, *et al*. Disruption of circadian rhythms accelerates development of diabetes through pancreatic beta-cell loss and dysfunction. *J Biol Rhythms*. 2011;**26**(5):423–433.
7. Haus E, *et al*. Biological clocks and shift work: circadian dysregulation and potential long-term effects. *Cancer Causes Control*. 2006;**17**(4):489–500.
8. Lévi F. Chronotherapeutics: the relevance of timing in cancer therapy. *Cancer Causes Control*. 2006;**17**(4):611–621.
9. Phillips AJK, *et al*. Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Sci Rep*. 2017;**7**(1):3216.
10. Buhr ED, *et al*. Molecular components of the Mammalian circadian clock. *Handb Exp Pharmacol*, 2013(217): 3–27.
11. Czeisler CA, *et al*. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science*. 1999;**284**(5423):2177–2181.
12. Bonmati-Carrion MA, *et al*. Protecting the melatonin rhythm through circadian healthy light exposure. *Int J Mol Sci*. 2014;**15**(12):23448–23500.
13. Czeisler CA, *et al*. Bright light induction of strong (type 0) resetting of the human circadian pacemaker. *Science*. 1989;**244**(4910):1328–1333.
14. Czeisler CA, *et al*. Entrainment of human circadian rhythms by light-dark cycles: a reassessment. *Photochem Photobiol*. 1981;**34**(2):239–247.
15. Duffy JF, *et al*. Entrainment of the human circadian system by light. *J Biol Rhythms*. 2005;**20**(4):326–338.
16. Jewett ME, *et al*. Revised limit cycle oscillator model of human circadian pacemaker. *J Biol Rhythms*. 1999;**14**(6):493–499.
17. Khalsa SB, *et al*. A phase response curve to single bright light pulses in human subjects. *J Physiol*. 2003;**549**(Pt 3):945–952.
18. Kronauer RE, *et al*. Quantifying human circadian pacemaker response to brief, extended, and repeated light stimuli over the phototopic range. *J Biol Rhythms*. 1999;**14**(6):500–515.
19. Forger DB, *et al*. A simpler model of the human circadian pacemaker. *J Biol Rhythms*. 1999;**14**(6):532–537.
20. Mistlberger RE, *et al*. Social influences on mammalian circadian rhythms: animal and human studies. *Biol Rev Camb Philos Soc*. 2004;**79**(3):533–556.
21. Mrosovsky N. Locomotor activity and non-photic influences on circadian clocks. *Biol Rev Camb Philos Soc*. 1996;**71**(3):343–372.
22. St Hilaire MA, *et al*. Addition of a non-photic component to a light-based mathematical model of the human circadian pacemaker. *J Theor Biol*. 2007;**247**(4):583–599.
23. Hannay KM, *et al*. Macroscopic models for human circadian rhythms. *J Biol Rhythms*. 2019;**34**(6):658–671.
24. Stone JE, *et al*. Application of a limit-cycle oscillator model for prediction of circadian phase in rotating night shift workers. *Sci Rep*. 2019;**9**(1):11032.

25. Woelders T, *et al*. Daily light exposure patterns reveal phase and period of the human circadian clock. *J Biol Rhythms*. 2017;**32**(3):274–286.
26. Martinez-Nicolas A, *et al*. Circadian monitoring as an aging predictor. *Sci Rep*. 2018;**8**(1):15027.
27. Zuurbier LA, *et al*. Fragmentation and stability of circadian activity rhythms predict mortality: the Rotterdam study. *Am J Epidemiol*. 2015;**181**(1):54–63.
28. Burgess HJ, *et al*. The dim light melatonin onset following fixed and free sleep schedules. *J Sleep Res*. 2005;**14**(3):229–237.
29. Crowley SJ, *et al*. A week in the life of full-time office workers: work day and weekend light exposure in summer and winter. *Appl Ergon*. 2015;**46 Pt A**:193–200.
30. Cheng P, *et al*. Shift work and cognitive flexibility: decomposing task performance. *J Biol Rhythms*. 2017;**32**(2): 143–153.
31. Cheng P, *et al*., Daytime sleep disturbance in night shift work and the role of PERIOD3. *J Clin Sleep Medicine*. 2018;**14**(3): 393–400.
32. Benloucif S, *et al*. Stability of melatonin and temperature as circadian phase markers and their relation to sleep times in humans. *J Biol Rhythms*. 2005;**20**(2):178–188.
33. Brown EN, *et al*. A mathematical model of diurnal variations in human plasma melatonin levels. *Am J Physiol*. 1997;**272**(3 Pt 1):E506–E516.
34. Crowley SJ, *et al*. Estimating dim light melatonin onset (DLMO) phase in adolescents using summer or school-year sleep/wake schedules. *Sleep*. 2006;**29**(12):1632–1641.
35. Kantermann T, *et al*. Comparing the Morningness-Eveningness Questionnaire and Munich ChronoType Questionnaire to the Dim Light Melatonin Onset. *J Biol Rhythms*. 2015;**30**(5):449–453.
36. Burgess HJ, *et al*. The relationship between the dim light melatonin onset and sleep on a regular schedule in young healthy adults. *Behav Sleep Med*. 2003;**1**(2):102–114.
37. Van Reen E, *et al*. Sex of college students moderates associations among bedtime, time in bed, and circadian phase angle. *J Biol Rhythms*. 2013;**28**(6):425–431.
38. Boivin DB, *et al*. Dose–response relationships for resetting of human circadian clock by light. *Nature*. 1996;**379**(6565):540–542.
39. Zeitzer JM, *et al*. Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *J Physiol*. 2000;**526 Pt 3**:695–702.
40. Wu G, *et al*. Population-level rhythms in human skin with implications for circadian medicine. *Proc Natl Acad Sci U S A*. 2018;**115**(48):12313–12318.
41. Wittenbrink N, *et al*. High-accuracy determination of internal circadian time from a single blood sample. *J Clin Invest*. 2018;**128**(9):3826–3839.
42. Mitchell JA, *et al*. Variation in actigraphy-estimated rest-activity patterns by demographic factors. *Chronobiol Int*. 2017;**34**(8):1042–1056.
43. Youngstedt SD, *et al*. Human circadian phase-response curves for exercise. *J Physiol*. 2019;**597**(8):2253–2268.
44. Joyce DS, *et al*. The accuracy of artificial and natural light measurements by actigraphs. *J Sleep Res*. 2020;**29**(5): e12963.
45. Stone JE, *et al*. Accuracy of the GENEActiv device for measuring light exposure in sleep and circadian research. *Clocks Sleep*. 2020;**2**(2):143–152.
46. Phillips AJK, *et al*. High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proc Natl Acad Sci U S A*. 2019;**116**(24):12019–12024.
47. St. Hilaire MA, *et al*. Prediction of individual differences in circadian adaptation to night work among older adults: application of a mathematical model using individual sleep–wake and light exposure data. *Chronobiol Int*. 2020;**37**(9-10):1404–1411.