SLEEPJ, 2021, 1–14

doi: 10.1093/sleep/zsab144
Advance Access Publication Date: 9 June 2021
Original Article

ORIGINAL ARTICLE

Optimal sleep and work schedules to maximize alertness

Francisco G. Vital-Lopez[1,2], Tracy J. Doty[3] and Jaques Reifman[1,*]

[1]Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Development Command, Fort Detrick, MD, USA, [2]The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD, USA and [3]Behavioral Biology Branch, Walter Reed Army Institute of Research, Silver Spring, MD, USA

*Corresponding author. Jaques Reifman, Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Development Command, ATTN: FCMR-TT, 504 Scott Street, Fort Detrick, MD 21702–5012, USA. Email: jaques.reifman.civ@mail.mil.

Abstract

Study Objectives: Working outside the conventional "9-to-5" shift may lead to reduced sleep and alertness impairment. Here, we developed an optimization algorithm to identify sleep and work schedules that minimize alertness impairment during work hours, while reducing impairment during non-work hours.

Methods: The optimization algorithm searches among a large number of possible sleep and work schedules and estimates their effectiveness in mitigating alertness impairment using the Unified Model of Performance (UMP). To this end, the UMP, and its extensions to estimate sleep latency and sleep duration, predicts the time course of alertness of each potential schedule and their physiological feasibility. We assessed the algorithm by simulating four experimental studies, where we compared alertness levels during work periods for sleep schedules proposed by the algorithm against those used in the studies. In addition, in one of the studies we assessed the algorithm's ability to simultaneously optimize sleep and work schedules.

Results: Using the same amount of sleep as in the studies but distributing it optimally, the sleep schedules proposed by the optimization algorithm reduced alertness impairment during work periods by an average of 29%. Similarly, simultaneously optimized sleep and work schedules, for a recovery period following a chronic sleep restriction challenge, accelerated the return to baseline levels by two days when compared to the conventional 9-to-5 work schedule.

Conclusions: Our work provides the first quantitative tool to optimize sleep and work schedules and extends the capabilities of existing fatigue-management tools.

Statement of Significance

Demands of a 24/7 society force millions of salary and wage earners to work outside of the conventional 9-to-5 shift, often leading to reduced sleep and alertness and compromising worker safety and productivity. These effects can be exacerbated or mitigated depending on the choice of when to sleep during rest opportunities or when to work, if afforded flexible work hours. Here, we describe an optimization algorithm that automatically identifies efficacious sleep and work schedules that result in minimum alertness impairment during work hours, while reducing impairment to the greatest extent possible during the rest of the day. This algorithm provides a new capability that complements existing resources for managing fatigue.

Key words: mathematical models; neurobehavioral performance; shift work; sleep deprivation

Submitted: 4 March, 2021; **Revised:** 26 May, 2021

Introduction

Approximately 30% of the U.S. labor force has work schedules outside of the conventional "9-to-5" day shift [1], precluding habitual night sleep on a regular basis. Such disruptions to the normal sleep-wake cycle are associated with negative health consequences as well as excessive fatigue and cognitive impairment [2–5]. In particular, cognitive impairment can be exacerbated or mitigated depending on the decision of when to sleep during available rest periods. However, this decision is not always obvious because the best time to sleep to enhance mental acuity may not be evident [6], or because social pressures and family obligations may dictate the sleep period. In addition, such a decision may also involve determining the best time to work, as nearly 60% of salary and wage earners in the United States have some flexibility in their work schedule [7].

A number of U.S. government agencies provide guidelines to help shift workers deal with extended and irregular shifts to better plan their rest-work schedules [8–10]. However, such guidelines are often too general. For example, for night-shift work, the U.S. Occupational Safety and Health Administration recommends between 7 and 9 h of daily sleep, with rest periods scheduled within 8 h of the start of the shift [8]. However, this guidance does not specify sleep start time or what to do if sleeping for at least 7 h is not possible. For nurses on shift work and long work hours, the National Institute for Occupational Safety and Health (NIOSH) recommends sleeping as long as possible before 1400, with preference to early morning time [10]. However, in some cases, it may not be physiologically possible to sleep for the needed duration. For instance, after a 24-h wakefulness period, an individual falling asleep around 0700, as recommended by the NIOSH guideline, may not be able to stay asleep for more than 4.5 h [11]. Hence, instead of using a "one-size-fits-all" approach, shift workers with fixed or varying work hours on regular or irregular work schedules would benefit from a publicly available evidence-based computer tool to help devise the most suitable rest-work schedule for each specific situation, while also considering social and family constraints.

A few commercial tools as well as a couple of publicly available ones are being marketed for predicting sleep episodes representative of a group of individuals. They are based on different mathematical models and are implemented in different computing platforms [12, 13], including Web servers [14–17], personal computers [14–16, 18–20], and smartphones [21, 22]. Given future work periods, these tools predict sleep episodes as well as subjective or objective alertness levels (or fatigue levels) associated with the provided schedules. While the detailed inner workings of such tools are often deemed to be proprietary information and, as such, are unavailable for peer review, in general, the predictions are based on one of three mathematical modeling approaches: the two-process model [23], the three-process model [24], or a statistical model [20], coupled with empirical constraints, such as historical activity patterns and commuting time before and after work. Surprisingly, with only one exception [21, 22], none of these tools attempts to identify the optimal time to sleep to maximize alertness during work periods. Rather, the predicted sleep episodes are simply an estimate of when sleep is most likely to occur, regardless of whether the predicted sleep episodes lead to peak alertness during the work periods. And, the one tool that maximizes alertness does so for a single point in time, for example, 0200 or 1000, rather than for the entire time duration of a work period [21, 22]. In addition, none of these tools attempts to simultaneously identify both the optimal time to sleep and the optimal time to work so that alertness is maximized during work periods. Rather, work periods are always fixed and provided as an input to the tool. Furthermore, none of these tools attempts to identify the optimal time to work that maximizes alertness for a fixed sleep schedule.

Planning effective sleep-work schedules requires three elements. First, we must have the ability to accurately predict the time course of alertness for any arbitrary sleep-wake schedule. This allows us to properly quantify alertness levels for any number of schedules, while accounting for potential constraints in sleep and work periods. Second, we must determine the extent to which an individual will be physiologically able to initiate and maintain sleep, depending on sleep history and time of day. Considering such physiological constraints is critical in determining the practicality and efficacy of a proposed sleep period. For example, as discussed above, planning for an 8-h sleep period starting at 0700 after 24 h of continuous wakefulness would be ineffective, given the inability to maintain sleep for more than ~4.5 h [11]. Third, we must have the ability to identify many feasible sleep-work schedules that meet sleep, work, and physiological constraints; compare and contrast them using a quantitative metric; and select the most effective one. To be most useful, this should be accomplished automatically, in real time, and instantiated in multiple computing platforms, including smartphones and publicly available Web servers [25].

Previously, we developed and extensively validated mathematical models that address the first and second elements mentioned above. Specifically, the Unified Model of Performance (UMP) predicts the effect of sleep history and time of day on alertness, as determined by the psychomotor vigilance test (PVT) [26, 27]. Recently, we extended the UMP to predict sleep latency (i.e. the timespan to fall asleep) and sleep duration, for an arbitrary sleep period [28]. To address the third element, here we describe an efficient optimization algorithm that searches the large number of possible sleep-work schedules and automatically provides in real time physiologically feasible schedules that maximize alertness during work periods, while satisfying sleep and work constraints. To this end, the algorithm identifies (1) optimal times to sleep for fixed work periods, (2) optimal times to work for fixed sleep periods, or (3) optimal combinations of sleep and work times, in each case leading to peak alertness during work periods. To assess the optimization algorithm, we first revisited three experimental sleep-deprivation studies [29–31] previously used to validate the UMP [26, 27, 32] and one additional study [33]. In these simulations, using the same total sleep time as in each of the original studies, we computed optimal sleep times that maximize alertness levels during fixed work periods. We then used the UMP to predict alertness impairment during the work periods using the original sleep schedules and compared these results against those obtained with the optimization algorithm. Finally, using the recovery phase of one of the studies, we simultaneously optimized sleep and work times and compared alertness impairment levels during work periods obtained with these optimal times against those of the original study.

Methods

Experimental studies

We used our algorithm to retrospectively optimize the sleep schedules of four diverse studies (*Studies 1–4*) involving chronic sleep restriction (CSR), total sleep deprivation (TSD) followed by CSR with daytime sleep, CSR with night-shift work, and CSR with day-night split sleep. For each study, the algorithm predicted sleep times that resulted in peak alertness during fixed (i.e. predefined) work periods. In addition, we simultaneously optimized sleep and work schedules during the recovery phase of *Study 1*, following 7 days of CSR. Below, we provide a brief description of each study.

Study 1 [31]. Twelve subjects [four men, mean age: 26 years, standard deviation (*SD*): 7.1 years] participated in a CSR and recovery study. Subjects were instructed to maintain their habitual sleep schedule at home for 14 days prior to start of the study. Their sleep patterns were monitored using actigraphy and sleep diaries. Then, subjects spent 8 nights in the laboratory, where they maintained their habitual sleep (mean duration: 7.1 h, *SD*: 0.7 h), waking up at 0700, and leaving the laboratory during the day to perform daily activities. Following this phase, subjects started the full-time, in-laboratory phase of the study, consisting of one baseline night of habitual sleep (waking up at 0700), seven nights of sleep restriction [time in bed (TIB) from 0400 to 0700], and five recovery nights (TIB from 2300 to 0700). Subjects performed a 5-min PVT every hour between 0800 and 1800 during the sleep restriction and recovery days. During PVT assessment and sleep periods, subjects stayed in individual sound-attenuated rooms, where the ambient temperature was maintained at 23°C and the lighting was kept at 500 lux during wake periods, while keeping background white noise at 65 dB at all times.

Study 2 [30]. Ten male Special Forces personnel (mean age: 28.6 years, range: 19 to 32 years) participated in a field study of sustained operations. After an overnight 8-h sleep period, starting at 0700 on day 1, the subjects underwent 31 h of TSD followed by 2 days of restricted daytime sleep (TIB from 1330 to 1730 on days 2 and 3). The study ended at 0930 on day 4. Subjects completed 31 sessions of a 5-min PVT.

Study 3 [33]. Twelve male subjects (mean age: 26.8 years, range: 18–32 years) participated in a study of simulated night-shift work. Subjects reported habitual sleep onset times between 2200 and 0200 and a total sleep time ranging from 6 to 9 h, prior to the in-laboratory simulated night-shift work. The simulated period started at 0700 on day 1 and ended at 2300 on day 5. Subjects had sleep opportunities with TIB from 0800 to 1200 on days 2 to 5 and performed 5-min PVTs every 1.5 h throughout most of the time awake. During PVT assessment and sleep periods, subjects stayed in individual sound-attenuated rooms, where the ambient temperature was maintained at 23°C and the lighting was kept at 500 lux during wake periods, while keeping background white noise at 65 dB at all times.

Study 4 [29]. Twelve male subjects (mean age: 28 years, range: 21–47 years) participated in a CSR study with split sleep. During the week prior to the study, subjects slept from 2330 to 0730, as verified by actigraphy and sleep diaries. Then, subjects started the full-time, in-laboratory phase of the study. After three baseline nights with TIB from 2330 to 0730, subjects started a period of 88 h (starting at 0700 on day 1 and ending at 2300 on day 4) in which they were allowed to sleep for 2 h every 12 h starting at 1500 on day 1 (seven TIB sleep opportunities in total). Subjects performed a 10-min PVT every 2 h throughout most of the time awake. Throughout the in-laboratory phase, subjects were isolated from time cues and ambient light was kept to less than 50 lux.

Metrics of alertness impairment

The PVT is a well-validated and widely used reaction-time test for assessing changes in alertness impairment levels caused by sleep loss [34, 35]. To perform a PVT, subjects are instructed to press a button (or tap on a touch screen) as quickly as possible immediately after they see a visual stimulus, which repeatedly appears on the screen at random time intervals between 1 and 10 s, over a typical 5- or 10-min PVT session. In essence, the PVT measures the response time (RT) from the presentation of a stimulus to a subject's response to it, from which we can compute a handful of statistics. Here, we use two PVT statistics: mean RT over the number of responses collected during a PVT session and PVT lapses, which indicate the number of RTs greater than 500 ms. Note that the larger the values of mean RT and PVT lapses are, the greater is the alertness impairment.

Multiple studies have demonstrated the equivalence between increases in alertness impairment level caused by sleep loss and increases in blood alcohol concentration (BAC) [36, 37]. Therefore, to provide a reference for undesirable values of mean RT and PVT lapses, we linked them to well-understood BAC limits of 0.06% and 0.08% (the federal limit to legally drive in the United States [38]). The alertness impairment caused by a BAC of 0.06% is equivalent to that caused by 19 h of wakefulness [25, 36, 37], which corresponds to a UMP-predicted alertness impairment level of 340 ms for mean RT (or seven lapses for PVT lapses). Similarly, the alertness impairment caused by a BAC of 0.08% is equivalent to that caused by 24 h of wakefulness, which corresponds to a UMP-predicted alertness impairment level of 460 ms for mean RT (or 12 lapses for PVT lapses). To assess the performance of the sleep/work optimization algorithm against the results obtained with the nominal sleep and work schedules used in the original studies, we computed the time duration for which the predicted alertness impairment was above the two BAC limits.

The UMP

Based on the seminal two-process model postulated by Borbély [23], we previously developed the UMP to quantitatively predict the effect of sleep history and time of day on alertness [26, 27]. The input to the UMP is a sleep-wake schedule, and the output is the corresponding prediction of the time course of the expected alertness impairment *P* representative of a group of individuals:

$$P(t) = S(t) + \kappa C(t), \tag{1}$$

where the homeostatic process *S* represents the need for sleep, process *C* represents the sleep-independent effect of the circadian rhythm on alertness, κ denotes the circadian amplitude, *t* represents time (in hours), and *P* is an estimate of the mean RT statistics or lapses of the PVT. We refer the reader to the Supplemental Material for the equations and parameters of the model.

We have extensively validated the UMP by comparing its predictions against data collected under various sleep conditions, including CSR (3–5 h of sleep per night for up to 7 days), TSD (28–88 h), combinations of TSD and CSR, daytime sleep, and sleep extension. Overall, we used data from 27 studies involving nearly 900 subjects (mostly healthy young adults). In particular, Ramakrishnan et al. [27] showed that 87% of the time, the UMP predictions were indistinguishable from experimental results.

Sleep-latency model

We recently extended the UMP to predict sleep latency $SL(t)$ at time t as function of P [28]:

$$SL(t) = A_{SL} e^{-k_{SL} P(t)} \quad (2)$$

where A_{SL} [272.4 min, standard error (SE): 58.5 min] represents a scaling factor and k_{SL} (0.012 ms^{-1}, SE: 8.1×10^{-4} ms^{-1}) denotes the rate at which $SL(t)$ decreases with P, which is computed using the UMP, for a given sleep-wake schedule.

Sleep-duration model

We also extended the UMP to predict sleep duration, assuming that sleep spontaneously ends when the homeostatic process S decreases to a circadian-regulated, sleep-termination threshold T representing the propensity of an individual to wake up [28]. The threshold $T(t)$ at time t is defined as:

$$T(t) = A_{SD} - \kappa_{SD} C(t + \varphi_{SD}) \quad (3)$$

where κ_{SD} (41.2 ms, SE: 3.8 ms) represents the amplitude of the threshold T, φ_{SD} (2.0 h, SE: 0.2 h) indicates a phase shift of the threshold T with respect to process C, and A_{SD} (211.6 ms) denotes a constant whose value is set so that process S reaches T at 0700 after sleep onset at 2300 under rested conditions.

We have validated these extensions to the UMP by comparing their predictions against data collected under various sleep conditions spanning the entire circadian cycle and different levels of sleep debt. Overall, we used data from 23 studies (309 subjects, mostly healthy adults), achieving average prediction errors of 4.0 min for sleep latency and 0.8 h for sleep duration [28].

Optimization algorithm

Previously, we developed an algorithm to identify optimal caffeine-dosing strategies to mitigate alertness impairment caused by sleep deprivation [39]. Following a similar approach, here we developed an algorithm to identify optimal sleep and work times that minimize alertness impairment (i.e. that maximize alertness) during work periods, while reducing impairment during non-work periods (i.e. the time an individual is awake but not working) to the greatest extent possible. Note that the algorithm can be applied to reduce alertness impairment in terms of PVT mean RT or PVT lapses. However, for simplicity, we provided a description of the algorithm using only the mean RT.

The algorithm's solutions also satisfy user-specified sleep and work constraints. In this formulation, a sleep schedule is comprised of a total of N sleep periods, where each sleep period n (with n = 1, 2, …, N) is defined by its start time S_n and duration D_n, as shown in Figure 1, A. Similarly, a work schedule is comprised of a total of M work periods, where each work period m (with m = 1, 2, …, M) is defined by its start time S_m and a fixed, user-defined duration D_m. By construction, the algorithm infers the total number of non-work periods Q based on the allocated sleep and work periods, where a non-work period q (with q = 1, 2, …, Q) is defined by its start time S_q and duration D_q. Accordingly, given fixed work duration D_m, the algorithm iteratively identifies optimal values for S_n, D_n, and S_m (and by extension for S_q and D_q) through three sequential steps, as described below and illustrated in Figure 2.

Algorithm inputs

The algorithm requires six inputs: (1) sleep window, that is, the time frame within which the algorithm can define sleep periods (Figure 1, A); (2) maximum sleep duration during each sleep window; (3) work window, that is, the time frame within which the algorithm can define work periods (Figure 1, A); (4) duration of the work period in each work window; (5) work alertness threshold, that is, the desired maximum impairment level during work periods (Figure 1, B); and (6) non-work alertness threshold, that is, the desired maximum impairment level during non-work periods (Figure 1, C). Note that sleep and work windows may overlap. Moreover, each sleep window may contain one or more sleep periods, as long as the combined sleep duration of all periods within a window does not exceed the maximum sleep duration for the sleep window. In contrast, we assumed that each work window contains only one continuous work period of fixed duration. Thus, N is variable and is determined by the algorithm, whereas M is fixed, as defined by the user-provided schedule. The alertness thresholds during work and non-work periods can be tuned to achieve the desired balance of alertness levels during wakefulness.

For a given set of inputs, the algorithm initially sets S_m of each work period m so that it coincides with the start of the work window and defines no sleep periods (i.e. it sets N = 0). We used these initial guesses for simplicity, as the algorithm is essentially insensitive to such selections. Iteratively, through sequential optimizations performed in each of three steps (Figure 2), the algorithm identifies a feasible solution that leads to peak alertness levels during work and non-work periods.

Optimization steps

In Step 1, for a given work period start time S_m, with m = 1, 2, …, M, the goal is to identify S_n and D_n for each sleep period n, with n = 1, 2, …, N, that minimize alertness impairment during the M work periods, as scored by the objective function Z_1 [Table 1, Equation (4)]. Figure 1, B graphically illustrates the variables used to compute Z_1. Z_1 has two terms to penalize sub-optimal solutions that result in alertness levels above the work alertness threshold for each work period m: the area under the UMP-predicted mean RT curve A_m and the maximum alertness impairment I_m. Then, we compute A_W as the sum of A_m (m = 1, 2, …, M) and I_W as the largest I_m (m = 1, 2, …, M) across all M work periods [Table 1, Equations (7) and (8)]. We normalized A_W and I_W by their corresponding values $A_{W,0}$ and $I_{W,0}$ for the no-sleep case (i.e. D_n = 0.0 for all n). Accordingly, the values for the objective function Z_1 range from zero (when the predicted alertness impairment is below the work alertness threshold for each of the M work periods) to 1.0 (when the predicted alertness impairment is equal to the one achieved with no sleep).

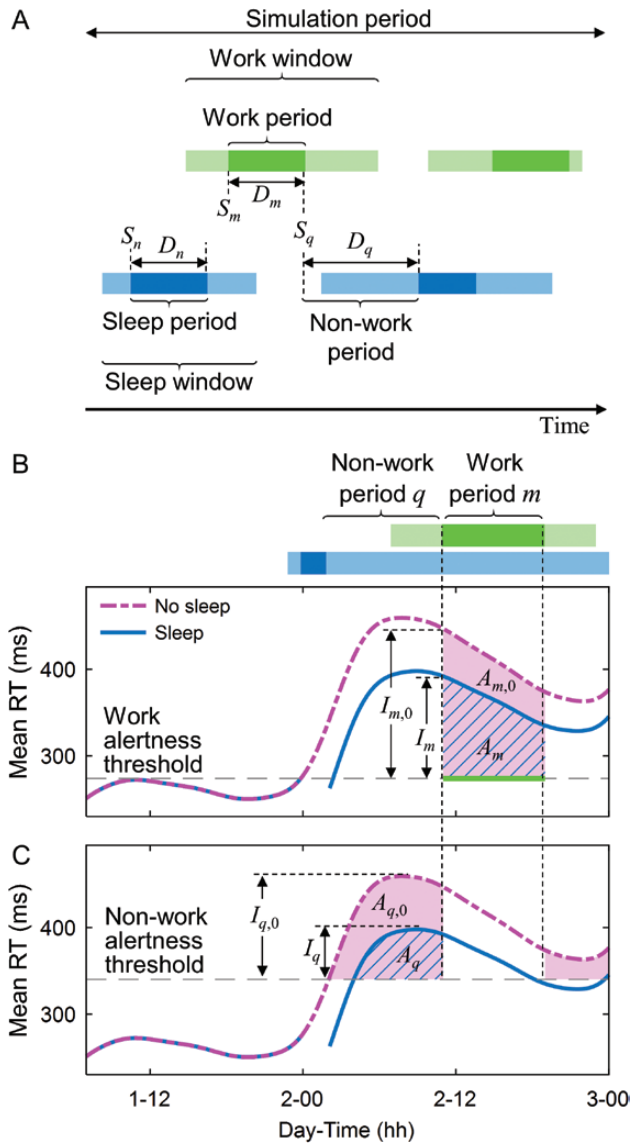

Figure 1. Definition of nomenclature and parameters used in the optimization algorithm. (A) A sleep schedule is comprised of sleep periods, where each sleep period n (dark blue rectangle) is defined by its start time S_n and duration D_n. A sleep window (light blue rectangle) defines the timespan for allocating sleep periods. Wakefulness is composed of work schedules and non-work schedules. A work schedule is comprised of work periods, where each work period m (dark green rectangle) is defined by its start time S_m and a fixed duration D_m specified by the user. A work window (light green rectangle) defines the timespan for allocating work periods. A non-work schedule is comprised of non-work periods when the individual is awake but not working, where each non-work period q is defined by its start time S_q and duration D_q, both indirectly inferred by the algorithm. (B) Quantities used to define the objective functions Z_1 and Z_3 in Table 1. The graph shows the Unified Model of Performance (UMP) predictions for the no-sleep condition (dashed pink line) and the current solution (blue line), indicated by the dark blue rectangle representing sleep at the top of the panel. For the work period m denoted by the dark green rectangle at the top of the panel, $A_{m,0}$ and A_m denote the corresponding areas, respectively, under the UMP predictions above the user-specified work alertness threshold (horizontal dash line). $I_{m,0}$ and I_m denote the difference between the peak of the UMP predictions for the no-sleep and sleep conditions, respectively, and the work alertness threshold for work period m. (C) Quantities used to define the objective function Z_2 in Table 1. The graph shows the two UMP predictions as in panel B. For the non-work period q immediately before work period m, $A_{q,0}$ and A_q denote the areas under the UMP predictions above the user-specified non-work alertness threshold (horizontal dash line). Likewise, $I_{q,0}$ and I_q denote the differences between the peak of the UMP predictions for the no-sleep and sleep conditions, respectively, and the non-work alertness threshold for non-work period q. RT: reaction time.

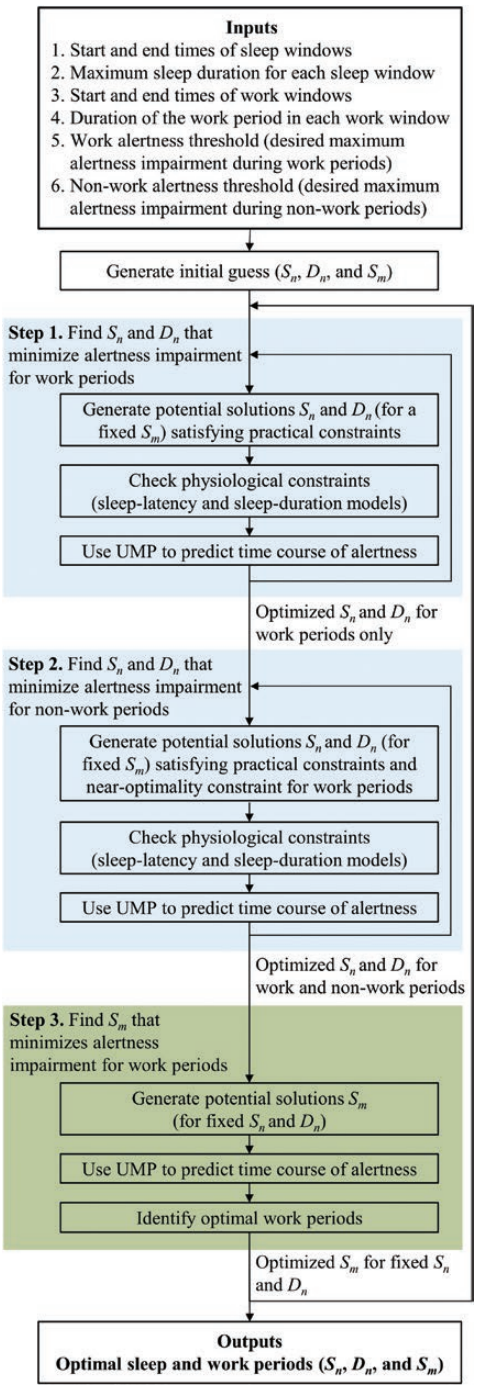

Figure 2. The three steps of the optimization algorithm to identify sleep and work schedules that minimize alertness impairment during work periods, while reducing impairment to the greatest extent possible during non-work periods. Throughout the algorithm, we used the Unified Model of Performance (UMP) and its extensions [26–28] to predict the time course of alertness and the physiological feasibility of each potential sleep schedule. S_n and D_n denote the start time and duration, respectively, of each sleep period n, and S_m denotes the start time of each work period m. Given the six user-provided inputs, the algorithm generates initial guesses for these three parameters. In Step 1, the algorithm identifies S_n and D_n that minimize alertness impairment during work periods, so as to meet a user-specified work alertness threshold. In Step 2, it updates S_n and D_n that minimize alertness impairment during non-work periods, so as to meet user-specified alertness thresholds for non-work periods while satisfying a near-optimality constraint for work periods [Table 1, Equation (19)]. Starting with these values of S_n and D_n, in Step 3, the algorithm identifies the optimal start times S_m for each work period m, so as to minimize alertness during work periods while meeting the user-defined threshold.

Table 1. Optimization algorithm objective functions and constraints used to identify the best sleep and work schedules that minimize alertness impairment during work periods, while reducing impairment to the greatest extent possible during non-work periods

Objective functions

$$\text{Step 1: } \min_{S_n, D_n} Z_1 = 0.5\left(\frac{A_W}{A_{W,0}} + \frac{I_W}{I_{W,0}}\right) \quad (4)$$

$$\text{Step 2: } \min_{S_n, D_n} Z_2 = 0.5\left(\frac{A_R}{A_{R,0}} + \frac{I_R}{I_{R,0}}\right) \quad (5)$$

$$\text{Step 3: } \min_{S_m} Z_3 = 0.5\left(\frac{A_W}{A_{W,0}} + \frac{I_W}{I_{W,0}}\right) \quad (6)$$

where Z_1, Z_2, and Z_3 denote the objective functions we wish to minimize in each of the three steps of the algorithm. S_n and D_n represent, respectively, the start time and duration (in hours) of sleep period n, with $n = 1, 2, \ldots, N$ (the total number of sleep periods), and S_m represents the start time (in hours) of work period m, with $m = 1, 2, \ldots, M$ (the total number of work periods). A_W, I_W, A_R, and I_R are defined as follows:

$$A_W = \sum_{m=1}^{M} A_m \quad (7)$$

$$I_W = \max\left(I_m,\ m = 1, \ldots,\ M\right) \quad (8)$$

$$A_R = \sum_{q=1}^{Q} A_q \quad (9)$$

$$I_R = \max\left(I_q,\ q = 1, \ldots,\ Q\right) \quad (10)$$

where A_m and A_q denote the area under the predicted psychomotor vigilance test mean response time (RT) curve above the alertness threshold for work period m and non-work period q, respectively, with $q = 1, 2, \ldots, Q$ (the total number of non-sleep periods). I_m and I_q denote the difference between the peak of the predicted mean RT curve and the threshold for work period m and non-work period q, respectively. $A_{W,0}$ and $A_{R,0}$ denote the total area under the mean RT curve (predicted assuming a no-sleep condition, that is, assuming $D_n = 0$ for $n = 1, 2, \ldots, N$), above the alertness threshold for the M work periods and the Q non-work periods, respectively. $I_{W,0}$ and $I_{R,0}$ denote the largest difference between the peak of the mean RT curve (predicted assuming no-sleep condition) and the alertness threshold across all M work periods and across all Q non-work periods, respectively.

Constraints

Allowed sleep durations

$$D_n \in \{0.5,\ 1.0,\ 1.5,\ 2.0,\ \ldots, 24.0\ \text{h}\} \quad (11)$$

$$S_n \in \{S_{n,\min}, S_{n,\min} + 0.5\ \text{h}, S_{n,\min} + 1.0\ \text{h}, S_{n,\min} + 1.5\ \text{h},\ \ldots, S_{n,\max} - D_n\} \quad (12)$$

where $S_{n,\min}$ and $S_{n,\max}$ denote the start and end times, respectively, of the sleep window containing sleep period n.

Allowed start time of each work period m

$$S_m \in \{S_{m,\min}, S_{m,\min} + 0.5\ \text{h}, S_{m,\min} + 1.0\ \text{h}, S_{m,\min} + 1.5\ \text{h},\ \ldots, S_{m,\max} - D_m\} \quad (13)$$

where $S_{m,\min}$ and $S_{m,\max}$ denote the start and end times, respectively, of the work window containing work period m and D_m denotes the duration of work period m.

Time between sleep periods

$$S_{n+1} - (S_n + D_n) \geq 6\ \text{h},\ \text{for}\, n \in \{1,\ 2,\ \ldots, N-1\} \quad (14)$$

Sleep periods should end at least 2 h before the following work period

$$S_m - (S_n + D_n) \geq 2\ \text{h},\ \text{for}\, m \in \{1,\ 2,\ \ldots, M\}\ \text{when}\, S_n \leq S_m \quad (15)$$

Sleep periods should start at least 2 h after the preceding work period

$$S_n - (S_m + D_m) \geq 2\ \text{h},\ \text{for}\, m \in \{1,\ 2,\ \ldots, M\}\ \text{when}\, S_n \geq S_m \quad (16)$$

Sleep duration should not exceed the predicted maximum sleep duration ($D_{n,\max}$)

$$D_n \leq D_{n,\max},\ \text{for}\, n \in \{1, 2,\ \ldots, N\} \quad (17)$$

Sleep duration should be at least 1 h if the predicted sleep latency (SL_n) is longer than 0.25 h

$$D_n \geq 1\ \text{h},\ \text{for each}\ n\ \text{that}\ SL_n \geq 0.25\ \text{h} \quad (18)$$

Maximum alertness impairment during work periods (Step 2 only)

$$A_m \leq A_{m,1} + \delta \times D_m,\ \text{for}\, m \in \{1,\ 2,\ \ldots, M\} \quad (19)$$

where $A_{m,1}$ denotes the area under the mean RT curve for work period m predicted with the current solution S_n and D_n obtained in Step 1, δ represents a tolerance threshold (δ = 30 ms), and D_m denotes the duration of work period m.

To obtain the solution for Step 1, the algorithm first identifies the wake period (i.e. the timespan between two consecutive sleep periods) with the worst alertness impairment across the work periods within each wake period (see Supplemental Material). Then, it assesses the benefit of eight potential changes to the current sleep schedule that are likely to reduce alertness impairment during the identified wake period, consequently reducing Z_1 (see Supplemental Material for details about the potential changes). The changes in sleep schedule must satisfy the following practical constraints defined in Table 1: (1) D_n must be a multiple of 0.5 h [Equation (11)]; 2) S_n and S_m should start on the hour or at the half-hour mark, for example, 2300 or 2330 [Equations (12) and (13)]; (3) the time lapse between sleep periods should be at least 6.0 h [Equation (14)]; and (4) the time between a sleep period and a work period should be at least 2.0 h [Equations (15) and (16)]. In addition, for each sleep period n, the potential solutions must satisfy two physiological constraints as defined in Table

1: (1) D_n cannot exceed the maximum sleep duration predicted by the sleep-duration model [Equation (17)] and (2) D_n should be at least 1.0 h if the sleep latency predicted by the sleep-latency model is greater than 15 min [Equation (18)]. The latter constraint is used to avoid scheduling a 30-min nap at a time when an individual would spend most of the sleep period trying to fall asleep. Note that when we predict alertness impairment corresponding to a given sleep schedule, the predicted sleep latency is not subtracted from the duration of the sleep period because the UMP was developed using time in bed as an input rather than total sleep time.

Starting with the solution obtained in Step 1, the goal of Step 2 is to modify S_n and D_n for each sleep period n such that alertness impairment across all Q non-work periods is minimized, as scored by the objective function Z_2 [Table 1, Equation (5)], while maintaining the impairment level across all M work periods near the optimal value achieved in Step 1. Figure 1, C graphically illustrates the variables used to compute the objective function Z_2. Similar to Z_1, Z_2 also has two terms to penalize suboptimal solutions for non-work periods that result in alertness levels above their threshold for each non-work period q: the area under the UMP-predicted mean RT curve A_q and the maximum alertness impairment I_q. Then, we compute A_R as the sum of A_q (q = 1, 2, …, Q) and I_R as the largest I_q (q = 1, 2, …, Q) across all M work periods [Table 1, Equations (9) and (10)]. We normalized A_R and I_R by their corresponding values $A_{R,0}$ and $I_{R,0}$ for the no-sleep case (i.e. D_n = 0.0 for all n). Accordingly, the values for the objective function Z_2 range from zero (when the predicted alertness is below the non-work alertness threshold for each of the Q non-work periods) to 1.0 (when the predicted alertness is equal to the one achieved with no sleep).

The procedure to minimize Z_2 is almost identical to the one used to minimize Z_1 in Step 1 (see Supplemental Material for details), except that here the potential solutions must also satisfy an additional constraint [Equation (19) in Table 1] to yield alertness impairment levels across all M work periods close to the one obtained in Step 1. This constraint ensures that the average impairment level during each work period m (with m = 1, 2, …, M) does not increase by more than a tolerance threshold δ. Here, we set δ = 30 ms, which is the estimated within-subject variability based on a 10-min PVT during well-rested conditions [40]. Thus, the outputs of Step 2 are the updated values of S_n and D_n that minimize alertness impairment during the non-work periods while maintaining near-optimal alertness impairment levels during work periods.

With the values obtained for S_n and D_n in Step 2, the goal of Step 3 is to determine the start times S_m for each work period m that minimize alertness impairment across all M work periods, as scored by the objective function Z_3 [Table 1, Equation (6)]. Z_3 is defined similarly as Z_1 (Table 1 and Figure 1, B). To obtain S_m, the algorithm generates two potential solutions for each work period m by adding or subtracting 0.5 h to S_m and selecting the solution that yields the largest reduction in Z_3. Note that the algorithm updates S_n to satisfy Equations (15) and (16) in Table 1. Then, the outputs of this step, that is, the updated S_m along with the current values for S_n and D_n, are used as inputs to Step 1 for the next iteration. We repeat this three-step procedure until a maximum number of iterations is reached or Z_3 cannot be further reduced. The final outputs are the optimal sleep schedules (S_n and D_n) and the optimal work schedules (S_m) that minimize alertness impairment during work periods, while reducing to the greatest possible extent the impairment during non-work periods.

Alertness impairment improvement metric

To quantify the benefit achieved by the optimization algorithm, we computed the difference between the areas under the UMP-predicted mean RT curve [A_W; Equation (7) in Table 1] for the original schedule used in the study and the corresponding optimal schedule, and expressed this difference as a percentage of the area for the original schedule.

Results

Optimizing sleep schedules

To assess the performance of the optimization algorithm for estimating the best times to sleep so as to minimize alertness impairment during wakefulness, we compared the algorithm's alertness predictions against those obtained with the original sleep schedules for each of the four experimental studies. However, because the optimal solutions are highly dependent on the accuracy of the UMP predictions, we first evaluated the model by comparing its alertness-impairment predictions against the experimental data of the original sleep schedules. Figures 3, A, 4, A, 5, A, and 6, A show the original sleep periods (orange rectangles above the graph), the measured group-average experimental data (black dots, representing either the mean RT in ms or the number of PVT lapses) and their associated two standard errors of the mean, and the group-average UMP predictions (dashed-dot orange line). Over the four studies, 78% of the UMP predictions (range: 73%–84%) fell within the two standard errors of the mean, indicating that the group-average predictions were generally indistinguishable from the experimentally measured mean values. Then, for each study, we used the optimization algorithm to identify the best sleep schedules that minimize alertness impairment during predetermined work periods and the ensuing non-work periods, while constraining the maximum sleep duration to the values used in each experimental study. Note that in these assessments, we only optimized the start time S_n and duration D_n of each sleep period n, while choosing to fix the start time of each work period S_m and their duration to a predetermined value. Thus, in these simulations, we did not involve Step 3 of the optimization algorithm.

For all studies, we set the work alertness threshold to 274 ms or four lapses, for mean RT and PVT lapses, respectively, which correspond to the highest predicted alertness impairment levels during wakefulness for recurring 8-h sleep periods between 2300 and 0700. Similarly, we set the non-work threshold to 340 ms for mean RT or seven lapses for PVT lapses, equivalent to the impairment caused by a BAC of 0.06% [36, 37]. Note that for *Study* 4, once we had obtained the optimal sleep schedule, we used the UMP to predict the alertness performance in terms of PVT lapses (see Supplemental Material for the UMP parameter values for PVT lapses), as this was the statistic reported in this study.

For *Study* 1, we simulated the first 5 days of evening shift work, where we assumed an 8-h work period from 1600 to 0000 (Figure 3, B, solid horizontal green line). We also assumed sleep windows of 12-h lengths, providing opportunities to sleep between shifts from 0200 to 1400 (Figure 3, A, light blue rectangles

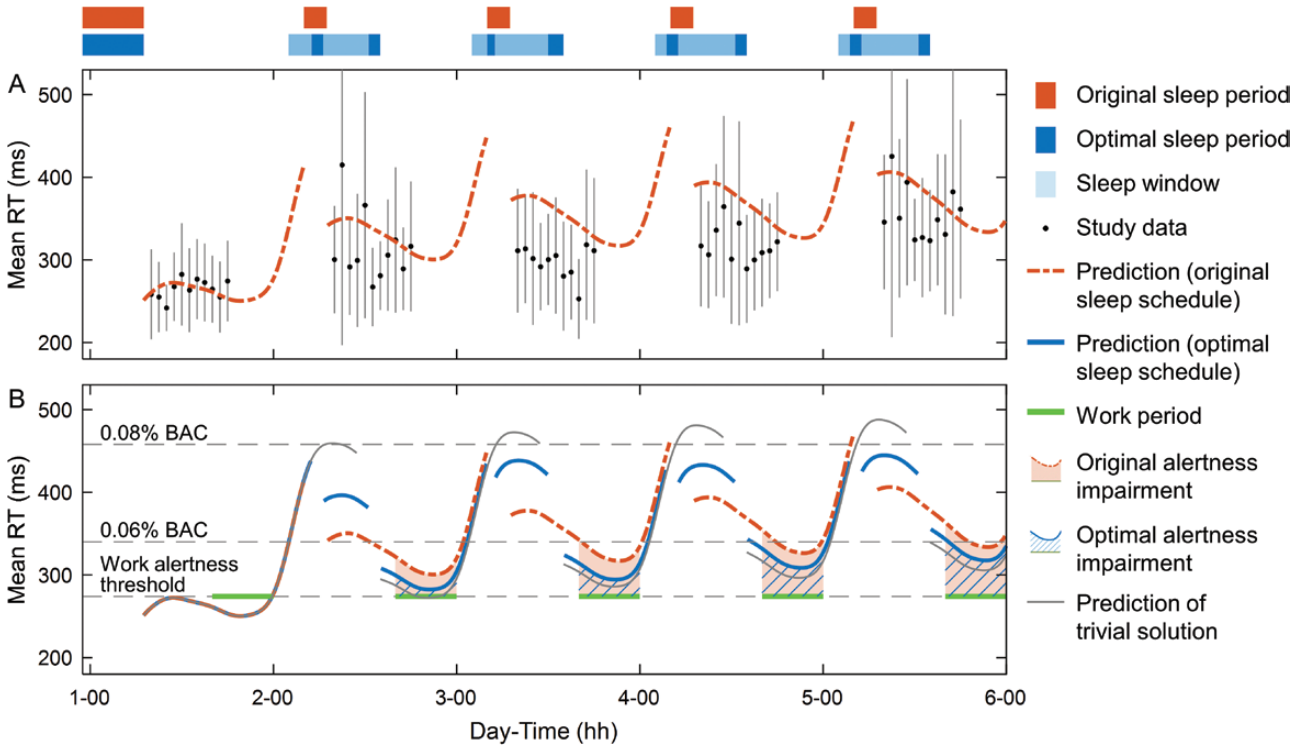

Figure 3. Optimal sleep schedule to minimize alertness impairment during work periods for *Study* 1. (A) Group-average values (12 subjects) of the experimentally measured mean reaction time (RT; black dots) for the psychomotor vigilance test, associated two standard errors of the mean, and Unified Model of Performance (UMP) predictions (dashed-dotted orange lines) for the original sleep schedule in the study (dark orange rectangles at the top of the panel). The top of the panel also shows the sleep windows (light blue rectangles) and the optimal sleep periods (dark blue rectangles). (B) UMP predictions for the original sleep schedule (dashed-dotted orange line) and the optimal sleep schedule (solid blue line). The orange-shaded and the blue-hatched areas under the mean RT curves indicate the alertness impairment above the work alertness threshold (274 ms, lower horizontal dashed line) during work periods (horizontal green line) for the original and optimal schedules, respectively. The upper (460 ms) and middle (340 ms) horizontal dashed lines indicate the alertness impairment equivalent to those of a blood alcohol concentration (BAC) of 0.08% and 0.06%, respectively. The graph also shows the UMP prediction for the "trivial" solution (solid thin gray line), where the 3-h sleep is scheduled from 1100 and 1400, immediately before each work period.

above the graph), with a maximum sleep duration for each sleep window set to 3.0 h, as in the original study. Figure 3, B shows the predicted mean RT for the original 0400 to 0700 nightly sleep schedule (dashed-dot orange line) and for the optimal sleep schedule (solid blue line). Interestingly, the optimal sleep schedule split the 3.0-h sleep into two bouts, one in the early morning hours, starting between 0330 and 0500 on different days, and one terminating at the end of the sleep window, that is, 2.0 h before the beginning of the evening shift (Figure 3, A, dark blue rectangles above the graph). Neither the original nor the optimal sleep schedule reduced alertness impairment during work periods below the desired threshold because the sleep duration was insufficient to achieve this goal. However, when compared to the original sleep schedule, by placing a sleep period immediately before the start of each shift, the optimal schedule reduced the average alertness impairment by 37% (Table 2; computed as the difference between the orange-shaded areas and the blue-hatched areas relative to the orange-shaded areas in Figure 3, B). Moreover, in the original schedule, the predicted alertness impairment across the 40.0-h work period was higher than the equivalent 0.06% BAC threshold for 16% of the time (i.e. for 6.5 h), while for the optimal schedule, it exceeded this level for only 0.5 h (Table 2). In addition, in the original schedule, the alertness impairment during non-work periods was higher than the equivalent 0.08% BAC threshold for 0.5 h, whereas in the optimal schedule, the impairment remained below this level the entire time.

In *Study* 2, we simulated three night shifts, where we assumed an 8-h work period from 0000 to 0800 (Figure 4, B, solid horizontal green line) and 12-h sleep windows from 1000 to 2200 preceding each shift (Figure 4, A, light blue rectangles above the graph), with a maximum sleep duration of 8.0 h, as in the original study. Figure 4, B shows the predicted mean RT for the original nightly sleep schedule (dashed-dot orange line) and for the optimal sleep schedule (solid blue line). The original schedule had two sleep periods from 1330 to 1730 (Figure 4, A, orange rectangles above the graph), whereas the optimal schedule identified three sleep periods (Figure 4, A, dark blue rectangles above the graph), one before the start of each work period, allocating a longer sleep time before the third work period (3.5 h) and a shorter sleep time before the first one (1.5 h). Compared to the original schedule, the optimal schedule reduced the average alertness impairment by 19% (Table 2; computed as the difference between the orange-shaded areas and the blue-hatched areas relative to the orange-shaded areas in Figure 4, B). Such a reduction is attributed primarily to the improvement in alertness impairment during the first night shift resulting from the preceding sleep period. The optimal schedule also kept impairment below the equivalent 0.08% BAC threshold for the entire work and non-work periods, while in the original schedule, alertness impairment exceeded this level for 1.3 h during work periods and 0.8 h during non-work periods.

In *Study* 3, we simulated a different version of a night shift, with five 8-h work periods from 2200 to 0600 (Figure 5, B, solid

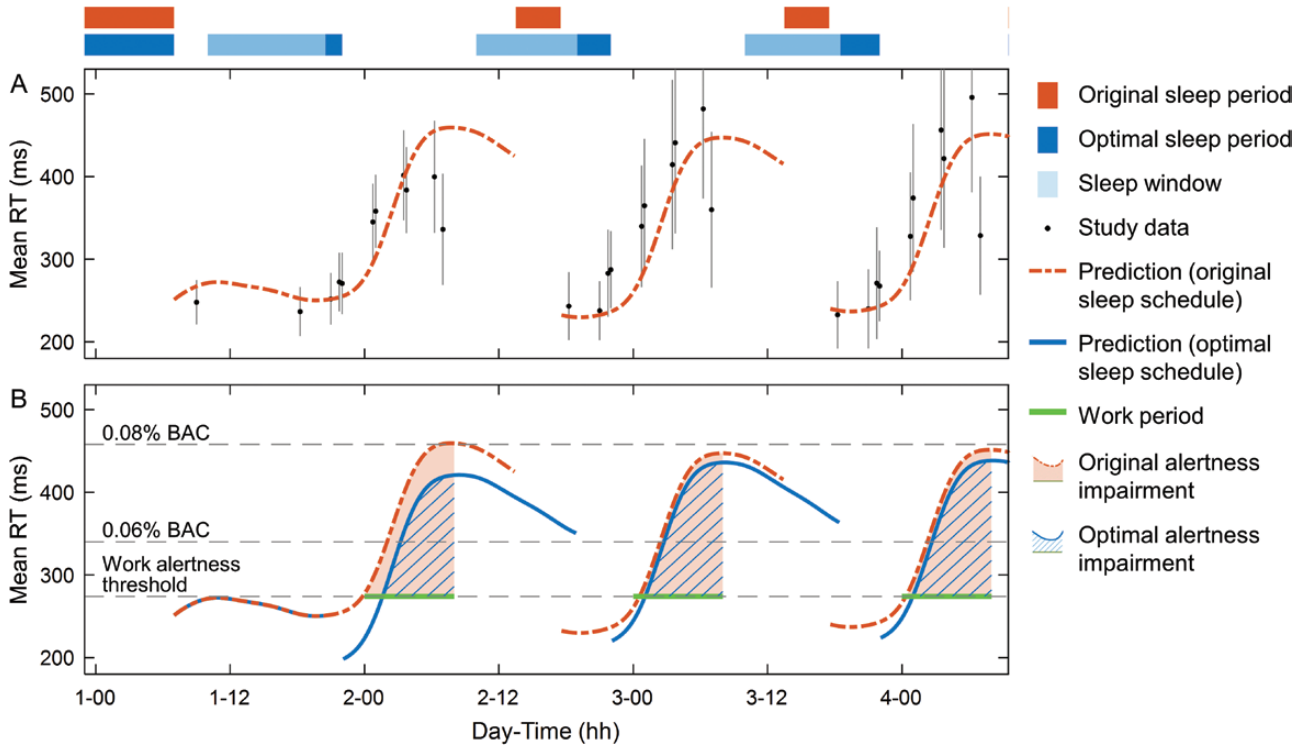

Figure 4. Optimal sleep schedule to minimize alertness impairment during work periods for *Study 2*. (A) Group-average values (10 subjects) of the experimentally measured mean reaction time (RT; black dots) for the psychomotor vigilance test, associated two standard errors of the mean, and Unified Model of Performance (UMP) predictions (dashed-dotted orange lines) for the original sleep schedule in the study (dark orange rectangles at the top of the panel). The top of the panel also shows the sleep windows (light blue rectangles) and the optimal sleep periods (dark blue rectangles). (B) UMP predictions for the original sleep schedule (dashed-dotted orange line) and the optimal sleep schedule (solid blue line). The orange-shaded and the blue-hatched areas under the mean RT curves indicate the alertness impairment above the work alertness threshold (274 ms, lower horizontal dashed line) during work periods (horizontal green line) for the original and optimal schedules, respectively. The upper (460 ms) and middle (340 ms) horizontal dashed lines indicate the alertness impairment equivalent to those of a blood alcohol concentration (BAC) of 0.08% and 0.06%, respectively.

green line) and 12-h sleep windows from 0800 to 2000 preceding each shift (Figure 5, A, light blue rectangles above the graph), with a total sleep duration of 16 h, as in the original study. Figure 5, B shows the predicted mean RT for the original sleep schedule (dashed-dot orange line) and the optimal sleep schedule (solid blue line). The original schedule had sleep periods from 0800 to 1200 (Figure 5, A, orange rectangles above the graph) after each shift. In contrast, the optimal schedule (Figure 5, A, dark blue rectangles above the graph) identified a 1.0-h sleep period before the first night shift, two sleep periods (one after a shift and the other before the next shift) during days 2 and 3, and one sleep period before the shifts on days 4 and 5 (the simulation period ended at 0800 on day 6). The optimal schedule reduced the average alertness impairment by 27% (Table 2; computed as the difference between the orange-shaded areas and the blue-hatched areas relative to the orange-shaded areas in Figure 5, B), mainly by scheduling a sleep period immediately before each shift, in contrast with the original schedule that included a 4-h sleep period after each shift. The optimal schedule sustained alertness impairment below the level equivalent to a BAC of 0.08% for the entire time, whereas in the original schedule alertness impairment was above this level for 4.0 h and 9.2 h during work and non-work periods, respectively. Importantly, most of the time, the impairment level of the original schedule exceeded the 0.08% BAC threshold right after the end of the shift, when the workers would be driving home.

In *Study 4*, we used the optimization algorithm to solve a variant of the previous problems, where the objective here was to identify the best way to distribute 14.0 h of sleep over an 88.0-h span so that alertness was optimal across the remaining 74.0 h of wakefulness (Figure 6, A). In this variation, we only used Step 1 of the algorithm because there were no non-work periods and the work periods equaled the entire wake time. In addition, we relaxed the constraints in Equations (15) and (16) in Table 1, allowing sleep periods to end immediately before work periods and to start immediately after work periods, respectively. Figure 6, B shows the predicted PVT lapses for the original sleep schedule (dashed-dot orange line) and for the optimal sleep schedule (solid blue line). While the original sleep schedule consisted of seven 2-h sleep periods, starting at 1500 or 0300 (Figure 6, A, orange rectangles above the graph), the optimal schedule had three consolidated sleep periods of either 4.5 or 5.0 h each, starting between 0100 and 0200 (Figure 6, A, dark blue rectangles above the graph). Compared to the original schedule, the optimal sleep schedule reduced the average alertness impairment by 31% (Figure 6, B and Table 2). This was achieved by the longer sleep periods during the night that more effectively reduced alertness impairment in the early morning hours during the most unfavorable circadian times. In contrast, the sleep periods starting at 1500 in the original schedule unnecessarily reduced impairment way below the desired threshold level during the late afternoon and early evening. Moreover, alertness impairment was above the equivalent 0.06% BAC threshold for 10.4 h in the optimal schedule vs. 27.5 h in the original schedule.

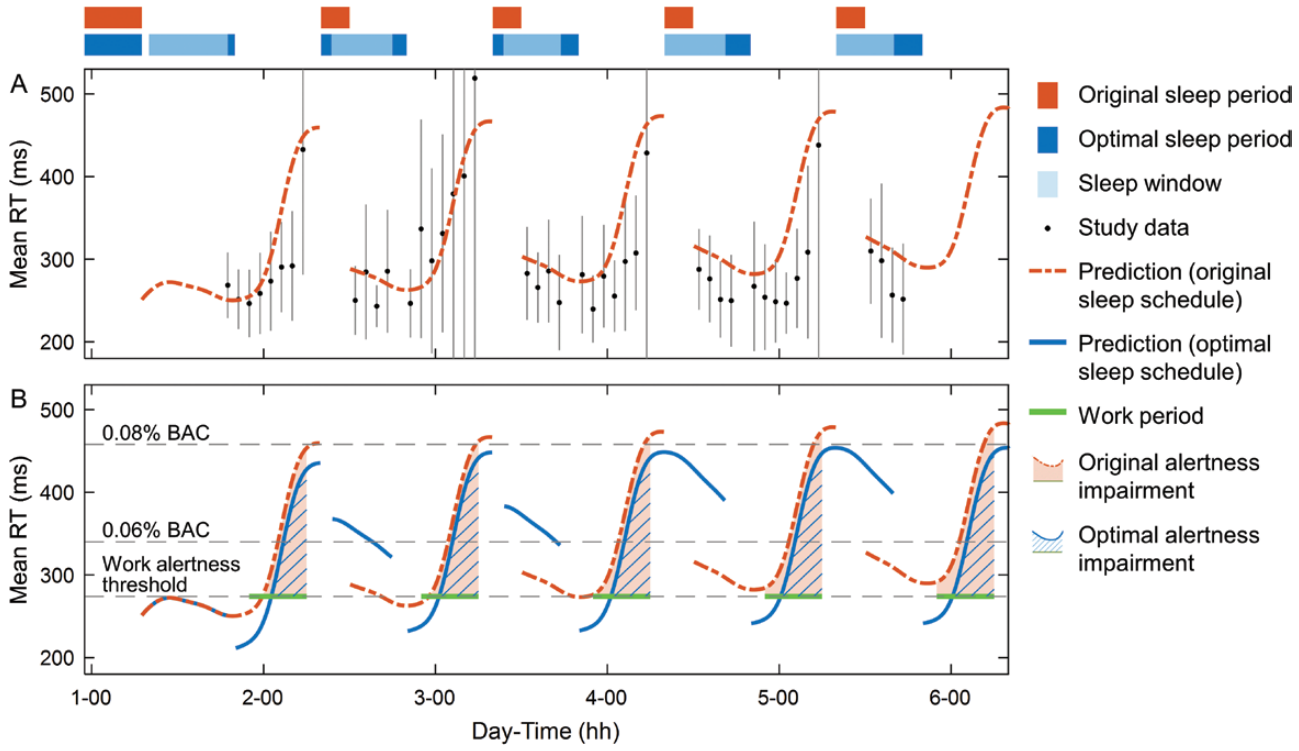

Figure 5. Optimal sleep schedule to minimize alertness impairment during work periods for *Study 3*. (A) Group-average values (12 subjects) of the experimentally measured mean reaction time (RT; black dots) for the psychomotor vigilance test, associated two standard errors of the mean, and Unified Model of Performance (UMP) predictions (dashed-dotted orange lines) for the original sleep schedule in the study (dark orange rectangles at the top of the panel). The top of the panel also shows the sleep windows (light blue rectangles) and the optimal sleep periods (dark blue rectangles). (B) UMP predictions for the original sleep schedule (dashed-dotted orange line) and the optimal sleep schedule (solid blue line). The orange-shaded and the blue-hatched areas under the mean RT curves indicate the alertness impairment above the work alertness threshold (274 ms, lower horizontal dashed line) during work periods (horizontal green line) for the original and optimal schedules, respectively. The upper (460 ms) and middle (340 ms) horizontal dashed lines indicate the alertness impairment equivalent to those of a blood alcohol concentration (BAC) of 0.08% and 0.06%, respectively.

Optimizing sleep and work schedules

To assess the performance of the optimization algorithm for simultaneously estimating the best times to sleep and the best times to work so as to minimize alertness impairment during wakefulness, we performed simulations for the 5 recovery days in *Study 1* that followed the 7 days of sleep restriction. Figure 7, A shows the original 2300 to 0700 sleep periods (orange rectangles above the graph), the measured group-average experimental data (black dots) and their associated two standard errors of the mean, and the group-average UMP predictions (dashed-dot orange line), which were indistinguishable from the experimental data. For the original study, we assumed a typical "9-to-5" work shift each day. For the optimization, we assumed daily 12-h sleep windows from 2100 to 0900, with a maximum sleep duration of 8 h per night, as in the original study, and a 12-h work window from 0700 to 1900, with an 8-h work duration. Using these inputs and the same alertness thresholds for work and non-work periods as described above, the algorithm identified optimal values for S_n, D_n, and S_m.

When compared to the original schedules, the optimal schedules were delayed by 2 h each day, that is, sleep periods from 0100 to 0900 (Figure 7, A) and work periods from 1100 to 1900 (Figure 7, B). This 2-h delay placed the work periods at a more favorable time in the circadian rhythm, resulting in a 26% reduction of the average alertness impairment during work (Figure 7, B) and a 75% reduction in the length of impairment above the 0.06% BAC (from 15.5 h in the original schedule to only 3.9 h). In addition, the optimal schedule resulted in a faster recovery, accelerating the return to baseline alertness levels by more than 2 days (Figure 7, C).

Discussion

Alertness impairment resulting from irregular shift work and the following reduction in sleep could be mitigated if sleep and work schedules were properly planned. Identification of the most suitable times to sleep and work requires the ability to screen a large number of schedules and assess their time course of alertness. Here, we developed an optimization algorithm to efficiently and effectively create and evaluate a very large number of physiologically feasible schedules and identify the ones that minimize alertness impairment during wakefulness. To this end, the algorithm relies on the well-validated UMP [26, 27] to predict the effect of each schedule on alertness impairment as well as on its recently developed model extensions to predict sleep latency and sleep duration [28] of each potential sleep period. The optimization algorithm solves three types of problems, each leading to peak alertness during work periods while reducing to the greatest possible extent alertness impairment during non-work periods. It identifies (1) optimal times to sleep for fixed (i.e. given) work periods, (2) optimal times to work for fixed sleep periods, or (3) optimal combinations of sleep and work times.

After demonstrating that the UMP properly predicted alertness impairment in four experimental studies that included

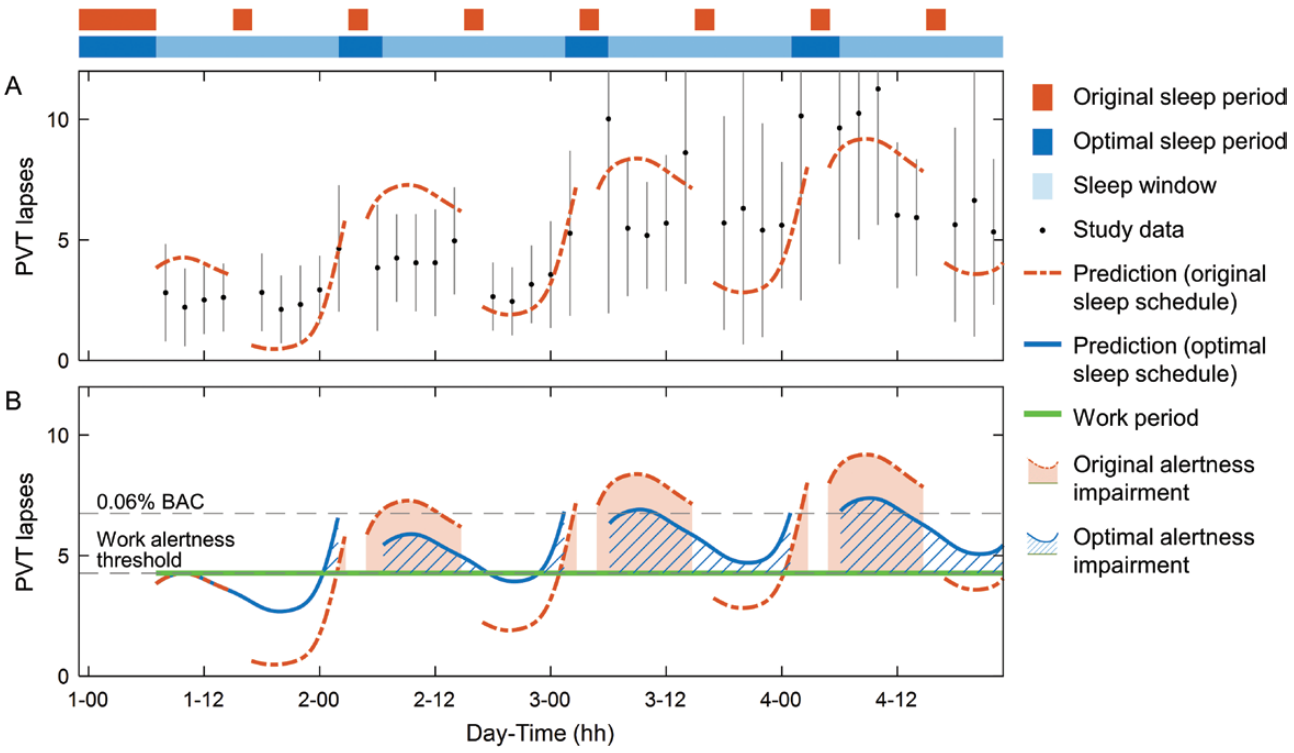

Figure 6. Optimal sleep schedule to minimize alertness impairment during work periods for *Study* 4. (A) Group-average values (12 subjects) of the experimentally measured lapses (black dots) for the psychomotor vigilance test (PVT lapses; black dots), associated two standard errors of the mean, and Unified Model of Performance (UMP) predictions (dashed-dotted orange lines) for the original sleep schedule in the study (dark orange rectangles at the top of the panel). The top of the panel also shows the sleep windows (light blue rectangles) and the optimal sleep periods (dark blue rectangles). (B) UMP predictions for the original sleep schedule (dashed-dotted orange line) and the optimal sleep schedule (solid blue line). The orange-shaded and the blue-hatched areas under the PVT lapses curves indicate the alertness impairment above the work alertness threshold (4 lapses, lower horizontal dashed line) during work periods (horizontal green line) for the original and optimal schedules, respectively. The upper horizontal dashed line at 7 lapses indicates the alertness impairment equivalent to that of a blood alcohol concentration (BAC) of 0.06%.

Table 2. Predicted average alertness impairment, percent improvement, and amount of time that the predicted alertness impairment for the original and optimal sleep schedules were above blood alcohol concentration (BAC) equivalents of 0.06% and 0.08% during the work periods

	Average alertness impairment (ms)			Time above 0.06% BAC equivalent (h)		Time above 0.08% BAC equivalent (h)	
Study	Original	Optimal	Improvement (%)	Original	Optimal	Original	Optimal
1	42.8	26.9	37	6.5	0.5	0.0	0.0
2	112.9	90.9	19	17.5	15.5	1.3	0.0
3	89.7	65.2	27	22.4	18.5	4.0	0.0
4*	1.6	1.1	31	27.5	10.4	0.0	0.0

*Average alertness impairment measured in lapses in the psychomotor vigilance test.

evening and night shifts as well as extended work periods, we assessed the optimization algorithm by comparing UMP-predicted alertness levels during wakefulness for sleep schedules proposed by the algorithm against those used in the original studies. Compared to the original sleep schedules, the optimal schedules reduced alertness impairment during work periods by 19%–37%—reductions that have practical implications. For example, the limited sleep opportunities in *Study* 4 yielded alertness impairments during wakefulness above the equivalent 0.06% BAC threshold for both sleep schedules (Figure 6, B). However, while the optimal schedule violated this threshold for 14% of the time (10.4 h out of the 74.0 h of wakefulness), the original sleep schedule surpassed the threshold for 27.5 h or 37% of the time. Similarly, while the optimal sleep schedules in *Studies* 2 and 3 never yielded alertness impairments above the equivalent 0.08% BAC level during work periods (Figures 4, B and 5, B), the original sleep schedules resulted in violations that lasted for 1.3 and 4.0 h, respectively. These results suggest that proper choices of suitable sleep schedules can have a considerable impact on alertness and, therefore, on worker's safety and productivity.

Scheduling available sleep opportunities immediately before work periods would invariably reduce alertness impairment during work. However, this simple strategy may sometimes result in undesirable impairment levels during non-work periods. For example, scheduling the 3-h sleep periods in *Study* 1 from 1100 to 1400, that is, immediately before each daily work period, would have resulted in impairment levels during non-work periods in days 2 to 5 that increasingly exceeded the equivalent impairment of a 0.08% BAC during the morning hours in each day (Figure 3, B, thin gray line). In fact, by day 5, the impairment

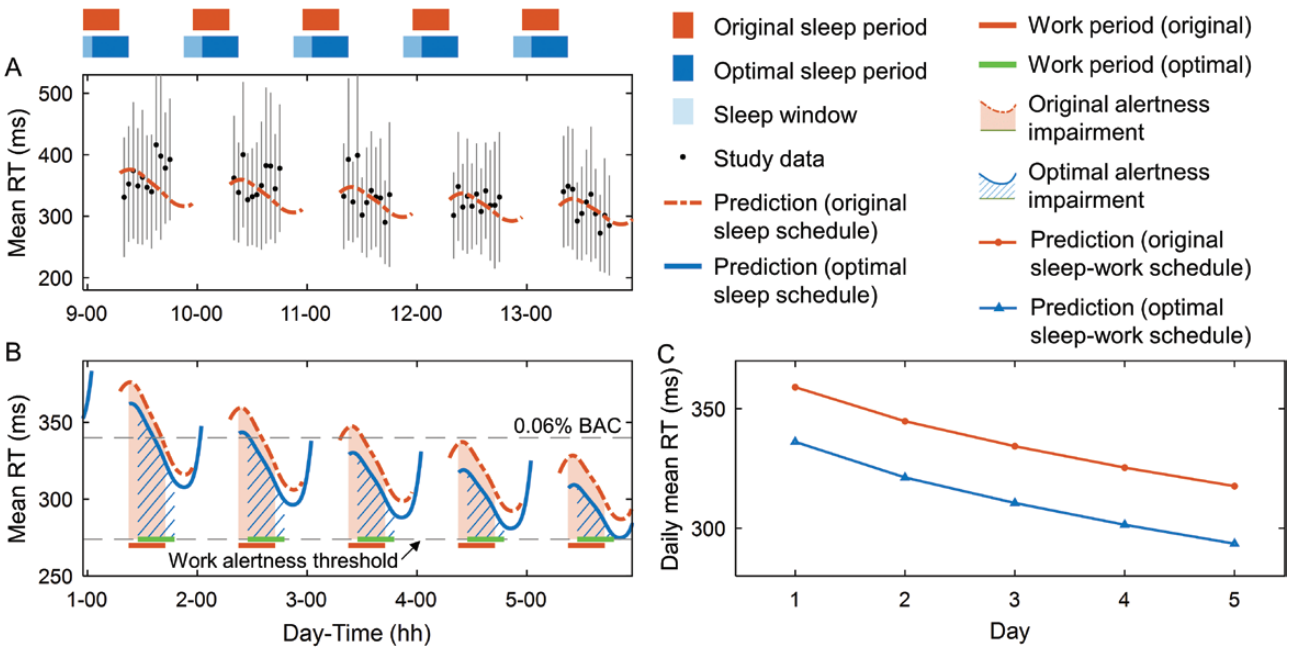

Figure 7. Optimal sleep and work schedules to minimize alertness impairment during work hours for the recovery period in *Study 1*, following 7 days of chronic sleep restriction (3 h of sleep per night). (A) Group-average values (12 subjects) of the experimentally measured mean reaction time (RT; black dots) for the psychomotor vigilance test, associated two standard errors of the mean, and Unified Model of Performance (UMP) predictions (dashed-dotted orange lines) for the original sleep schedule of the 5 recovery days (dark orange rectangles at the top of the panel). The top of the panel also shows the sleep windows (light blue rectangles) and the optimal sleep periods (dark blue rectangles). (B) UMP predictions for the original sleep schedule (dashed-dotted orange line) and the optimal sleep schedule (solid blue line). The horizontal orange and green solid lines represent the original (0900–1700) and optimal (1100–1900) work schedules, respectively, and the orange-shaded and blue-hatched areas represent the alertness impairment during the corresponding work period. The upper horizontal dashed line at 340 ms indicates the alertness impairment equivalent to that of a blood alcohol concentration (BAC) of 0.06%. (C) Daily mean RT (averaged during the work period) as a function of recovery days.

would have exceeded this limit during more than 6.5 h. In contrast, by allocating about half of the total sleep time just before the work periods and the other half during the early morning hours when impairment was highest, the optimal schedule yielded impairment levels for non-work periods that remained below this level each day (Figure 3, B, solid blue line), while hardly increasing impairment during the work periods (i.e. <10 ms on average). This result illustrates that devising efficient sleep schedules is not necessarily a trivial task.

The simulation results also showed that small variations of very similar types of work schedules may require different optimal sleep times. For example, the night shifts in *Studies 2* and *3* differed only by the number of shifts (3 vs. 5) and their starting times (0000 vs. 2200), with the latter allowing for a small increase in the average daily sleep (0.5 h). In both studies, the optimization algorithm scheduled one sleep period just before each night shift (Figures 4 and 5); however, in *Study 3*, it also scheduled a sleep period after the shifts on days 2 and 3, indicating that a "one-size-fits-all" rule may result in suboptimal alertness levels.

For workers who have flexible work hours, optimizing sleep and work schedules simultaneously could provide further benefits. For example, the results in Figure 7, B showed that during the recovery period immediately after 7 days of CSR, delaying both sleep and work starting times by only 2 h with respect to the typical sleep (2300–0700) and work (0900–1700) schedules yielded a 26% reduction in alertness impairment during work hours (solid blue line vs. dashed-dotted orange line). The figure also suggests that a further delay in the work schedule would have resulted in even larger reductions in impairment. In this simple example, an experienced sleep scientist would have arrived at the same conclusion. However, the algorithm can provide optimal sleep and work schedules for scenarios where the proper solution may not be as apparent and for which quantitative predictions of the time course of alertness may be required to identify optimal schedules.

The proposed work has limitations. First, we developed the underlying models to predict the effect of sleep schedule on alertness, sleep latency, and sleep duration using study data from young, healthy adults. It is unknown the extent to which such predictions can be applied to older populations or populations with sleep disorders. Second, the models provided group-average predictions and did not account for individual variability, such as the level of resilience or vulnerability to sleep deprivation. However, this limitation can be mostly overcome by customizing the UMP predictions for each individual [41, 42]. Third, the UMP predicted the time course of alertness as determined by the PVT, a reaction time test. Therefore, the resulting optimal schedules may not necessarily optimize other aspects of cognitive performance. Fourth, only *Study 1* included female subjects (8 vs. 4 male subjects). Hence, the conclusions resulting from *Studies 2–4* are limited to male subjects. However, because these results are based on the UMP and its extensions [28], which we developed and validated using more than 31 distinct studies, of which 20 included a total of 270 female subjects (vs. a total of 553 male subjects), we believe that the results presented here are likely to be valid for women, as well. This is supported by the fact that the results in *Studies 2–4* are consistent with those in *Study 1* (Table 2). Finally, because the amount of sleep is finite, attempting to achieve peak alertness levels during work periods may increase alertness impairment during non-work periods (Figure 5, B). However, our algorithm allows the user to modulate this tradeoff by adjusting the alertness threshold levels for work and non-work periods.

In summary, here we present the first computational algorithm to optimize sleep and work schedules so as to maximize alertness during work and non-work hours. This unique capability complements other fatigue-management tools based on the UMP, such as the publicly available *2B-Alert* Web [25], and can be combined with our caffeine-consumption optimization algorithm [39] to provide a more comprehensive set of countermeasure strategies to mitigate alertness impairment due to limited sleep.

Supplementary material

Supplementary material is available at *SLEEP* online.

Funding

This work was sponsored by the Military Operational Medicine Research Program of the U.S. Army Medical Research and Development Command, Fort Detrick, MD.

Disclosure Statements

Financial disclosure: This was not an industry-supported study. F.G.V.L. and J.R. received royalties for the licensing of the *2B-Alert* technology.

Non-financial disclosure: The authors have indicated no conflicts of interest. The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army, the U.S. DoD, or The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. The investigators have adhered to the policies for protection of human subjects as prescribed in AR 70–25. This paper has been approved for public release with unlimited distribution.

References

1. Alterman T, *et al*. Prevalence rates of work organization characteristics among workers in the U.S.: data from the 2010 National Health Interview Survey. *Am J Ind Med*. 2013;**56**(6):647–659.
2. Akerstedt T. Sleepiness as a consequence of shift work. *Sleep*. 1988;**11**(1):17–34.
3. Lim J, *et al*. A meta-analysis of the impact of short-term sleep deprivation on cognitive variables. *Psychol Bull*. 2010;**136**(3):375–389.
4. Sallinen M, *et al*. Shift work, sleep, and sleepiness—differences between shift schedules and systems. *Scand J Work Environ Health*. 2010;**36**(2):121–133.
5. Wright KP Jr, *et al*. Shift work and the assessment and management of shift work disorder (SWD). *Sleep Med Rev*. 2013;**17**(1):41–54.
6. Capaldi VF, *et al*. Optimizing sleep in the military: challenges and opportunities. *Chest*. 2019;**155**(1):215–226.
7. U.S. Bureau of Labor Statistics. Job flexibilities and work schedules summary [press release]. September 24 2019. USDL-19-1691. https://www.bls.gov/news.release/flex2.nr0.htm. Accessed March 4, 2021.
8. U.S. Occupational Safety and Health Administration. Long work hours, extended or irregular shifts, and worker fatigue. https://www.osha.gov/SLTC/workerfatigue/prevention.html. Accessed March 4, 2021.
9. Army Techniques Publication 6–22.5. *A Leader's Guide to Health and Fitness*. Washington, DC: Headquarters, Department of the Army. 2016.
10. Caruso CC, *et al*. NIOSH training for nurses on shift work and long work hours. 2015. https://doi.org/10.26616/NIOSHPUB2015115revised042020. Accessed March 4, 2021.
11. Akerstedt T, *et al*. The circadian variation of experimentally displaced sleep. *Sleep*. 1981;**4**(2):159–169.
12. Branford K, *et al*. *Biomathematical Fatigue Models Guidance Document*. Australia: Civil Aviation Safety Authority (CASA); 2014.
13. Tucker P, *et al*. *Working Time, Health and Safety: A Research Synthesis Paper*. Geneva: International Labor Office; 2012.
14. Hursh S, *et al*. *Validation and Calibration of a Fatigue Assessment Tool for Railroad Work Schedules, Final Report*. Washington, DC: U.S. Department of Transportation; 2008.
15. Jeppesen. The Boeing alertness model technical fact sheet. 2019. https://ww2.jeppesen.com/wp-content/uploads/2019/06/BAM_Tech_Fact_Sheet_2.3.pdf. Accessed March 4, 2021.
16. McCauley P, *et al*. Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. *Sleep*. 2013;**36**(12):1987–1997.
17. Moore-Ede M, *et al*. Circadian alertness simulator for fatigue risk assessment in transportation: application to reduce frequency and severity of truck accidents. *Aviat Space Environ Med*. 2004;**75**(3 Suppl):A107–A118.
18. Belyavin AJ, *et al*. Modeling performance and alertness: the QinetiQ approach. *Aviat Space Environ Med*. 2004;**75**(3 Suppl):A93–103; discussion 104.
19. Dawson D, *et al*. A quantitative model of work-related fatigue: background and definition. *Ergonomics*. 2001;**44**(2):144–163.
20. Folkard S, *et al*. A Fatigue/Risk Index to assess work schedules. *Somnologie - Schlafforschung Schlafmedizin*. 2007;**11**(3):177–185.
21. Jeppesen. Jeppesen CrewAlert for iOS. 2019. https://ww2.jeppesen.com/wp-content/uploads/2019/05/CrewAlert-product-sheet.pdf. Accessed March 4, 2021.
22. Jeppesen. CrewAlert fatigue mitigation functionality. 2019. https://ww2.jeppesen.com/wp-content/uploads/2019/07/CrewAlert-Fatigue-Mitigation-Functionality.pdf. Accessed March 4, 2021.
23. Borbély AA. A two process model of sleep regulation. *Hum Neurobiol*. 1982;**1**(3):195–204.
24. Akerstedt T, *et al*. Validation of the S and C components of the three-process model of alertness regulation. *Sleep*. 1995;**18**(1):1–6.
25. Reifman J, *et al*. 2B-Alert web: an open-access tool for predicting the effects of sleep/wake schedules and caffeine consumption on neurobehavioral performance. *Sleep*. 2016;**39**(12):2157–2159.
26. Rajdev P, *et al*. A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. *J Theor Biol*. 2013;**331**:66–77.
27. Ramakrishnan S, *et al*. A unified model of performance: validation of its predictions across different sleep/wake schedules. *Sleep*. 2016;**39**(1):249–262.
28. Vital-Lopez FG, *et al*. Models for predicting sleep latency and sleep duration. *Sleep*. 2021;**44**(5):1–13.

29. Doran SM, *et al*. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch Ital Biol*. 2001;**139**(3):253–267.
30. Kamimori GH, *et al*. Caffeine improves reaction time, vigilance and logical reasoning during extended periods with restricted opportunities for sleep. *Psychopharmacology (Berl)*. 2015;**232**(12):2031–2042.
31. Rupp TL, *et al*. Banking sleep: realization of benefits during subsequent sleep restriction and recovery. *Sleep*. 2009;**32**(3):311–321.
32. Ramakrishnan S, *et al*. A unified model of performance for predicting the effects of sleep and caffeine. *Sleep*. 2016;**39**(10):1827–1841.
33. Wesensten NJ, *et al*. Ampakine (CX717) effects on performance and alertness during simulated night shift work. *Aviat Space Environ Med*. 2007;**78**(10):937–943.
34. Dinges DF, *et al*. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav Res Meth Ins C*. 1985;**17**:652–655.
35. Dorrian J, Rogers NL, Dinges DF. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: Kushida CA, ed. *Sleep Deprivation. Clinical Issues, Pharmacology, and Sleep Loss Effects*. New York: Marcel Dekker; 2005:39–70.
36. Dawson D, *et al*. Fatigue, alcohol and performance impairment. *Nature*. 1997;**388**(6639):235.
37. Williamson AM, *et al*. Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accid Anal Prev*. 2001;**33**(3): 313–326.
38. U.S. Alcohol Policy Information System. Blood alcohol concentration (BAC) limits. https://alcoholpolicy.niaaa.nih.gov/. Accessed March 4, 2021.
39. Vital-Lopez FG, *et al*. Caffeine dosing strategies to optimize alertness during sleep loss. *J Sleep Res*. 2018;**27**(5):e12711.
40. Rupp TL, *et al*. Trait-like vulnerability to total and partial sleep loss. *Sleep*. 2012;**35**(8):1163–1172.
41. Liu J, *et al*. Real-time individualization of the unified model of performance. *J Sleep Res*. 2017;**26**(6):820–831.
42. Reifman J, *et al*. 2B-Alert App: a mobile application for real-time individualized prediction of alertness. *J Sleep Res*. 2019;**28**(2):e12725.