



## REGULAR RESEARCH PAPER



# 2B-Alert App: A mobile application for real-time individualized prediction of alertness

Jaques Reifman<sup>1</sup> | Sridhar Ramakrishnan<sup>1</sup> | Jianbo Liu<sup>1</sup> | Adam Kapela<sup>1</sup> | Tracy J. Doty<sup>2</sup> | Thomas J. Balkin<sup>2</sup> | Kamal Kumar<sup>1</sup> | Maxim Y. Khitrov<sup>1</sup>

<sup>1</sup>Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, Maryland

<sup>2</sup>Behavioral Biology Branch, Walter Reed Army Institute of Research, Silver Spring, Maryland,

## Correspondence

Jaques Reifman, Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702.  
Email: jaques.reifman.civ@mail.mil

## Funding information

Medical Research and Materiel Command, Grant/Award Number: Military Operational Medicine Research Program

## Abstract

Knowing how an individual responds to sleep deprivation is a requirement for developing personalized fatigue management strategies. Here we describe and validate the 2B-Alert App, the first mobile application that progressively learns an individual's trait-like response to sleep deprivation in real time, to generate increasingly more accurate individualized predictions of alertness. We incorporated a Bayesian learning algorithm within the validated Unified Model of Performance to automatically and gradually adapt the model parameters to an individual after each psychomotor vigilance test. We implemented the resulting model and the psychomotor vigilance test as a smartphone application (2B-Alert App), and prospectively validated its performance in a 62-hr total sleep deprivation study in which 21 participants used the app to perform psychomotor vigilance tests every 3 hr and obtain real-time individualized predictions after each test. The temporal profiles of reaction times on the app-conducted psychomotor vigilance tests were well correlated with and as sensitive as those obtained with a previously characterized psychomotor vigilance test device. The app progressively learned each individual's trait-like response to sleep deprivation throughout the study, yielding increasingly more accurate predictions of alertness for the last 24 hr of total sleep deprivation as the number of psychomotor vigilance tests increased. After only 12 psychomotor vigilance tests, the accuracy of the model predictions was comparable to the peak accuracy obtained using all psychomotor vigilance tests. With the ability to make real-time individualized predictions of the effects of sleep deprivation on future alertness, the 2B-Alert App can be used to tailor personalized fatigue management strategies, facilitating self-management of alertness and safety in operational and non-operational settings.

## KEYWORDS

alertness, caffeine, individualized predictions, psychomotor vigilance test, sleep, smartphone app

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. *Journal of Sleep Research* published by John Wiley & Sons Ltd on behalf of European Sleep Research Society

## 1 | INTRODUCTION

Consumer-level, personal fitness devices allow us to continually track sleep time with reasonable accuracy in free-living conditions (Ferguson, et al., 2015). Likewise, ubiquitous mobile computing devices allow us to objectively measure alertness impairment due to sleep deprivation via a psychomotor vigilance test (PVT; Brunet, et al., 2017; Grant, et al., 2017). Although such data provide the means to quantify longitudinal patterns of sleep and alertness, the extent to which this knowledge results in *actionable* information is much less clear. Ideally, such devices should provide customized guidance to allow each sleep-deprived individual to maximize alertness at the desired times of the day for the desired duration.

Achieving this goal requires the ability to “learn” an individual’s trait-like response to sleep deprivation, and the ability to accurately and quantitatively forecast—at an individual level—the effects of sleep interventions, while considering daily variations in alertness due to time of day and previous schedules of sleep and caffeine consumption. To this end, recently, our group developed the Unified Model of Performance (UMP), a biomathematical model that predicts PVT performance under sleep-loss conditions ranging from chronic sleep restriction (CSR) to total sleep deprivation (TSD), while accounting for the restorative effects of caffeine (Ramakrishnan et al., 2014; Ramakrishnan, Wesensten, Balkin, et al., 2016; Ramakrishnan, Wesensten, Kamimori, et al., 2016). The UMP is well validated: we have assessed its predictions on 442 subjects from 14 different studies, including 24 different sleep/wake conditions (from 3 to 10 hr of sleep per night to 88 hr of TSD) and nine different caffeine conditions during sleep deprivation (from single/repeated doses of 100–600 mg). We have also shown that, when individually customized, the UMP captures the individual’s trait-like response to sleep deprivation under different sleep-deprivation challenges (Ramakrishnan et al., 2015).

More recently, through retrospective computer simulations mimicking real-time operation, we demonstrated that our validated UMP could automatically and continually learn an individual’s trait-like response to TSD and CSR challenges on the fly (Liu, et al., 2017). We showed that a Bayesian learning algorithm could progressively adapt the UMP parameters after each PVT, so that model predictions increasingly matched an individual’s sequence of PVT measurements during sleep restriction.

However, until now, these results were all based on retrospective analyses of simulated real-time performance, because we lacked a platform and algorithms for prospective, real-time operation. Here we aim to describe and validate the *2B-Alert* App, the first mobile app that uses the results of each PVT to automatically and progressively learn an individual’s trait-like response to sleep deprivation and to make individualized predictions of alertness in real time. To validate the app, we used it to collect PVT data before, during and after TSD in a laboratory-controlled study, and to make real-time prospective predictions of alertness during the TSD challenge.

## 2 | MATERIALS AND METHODS

### 2.1 | The *2B-Alert* App

Figure 1 shows screen-capture images of the major *2B-Alert* graphical user interfaces, including those for the main menu, sleep and caffeine schedules, sleep input, PVT session, and prediction displays.

### 2.2 | Inputs and outputs

User inputs to the *2B-Alert* App include: (a) sleep schedule (Figure 1b, c); (b) caffeine schedule; and (c) PVT data, which are obtained via tests administered within the app itself (Figure 1d). Sleep and caffeine schedules can be entered retroactively or proactively to explore the effects of prior or future interventions on alertness, respectively, and can be edited at any time. A wrist-worn device can also be used to automatically enter sleep data via actigraphy (Supporting Information Data S1, Section I). PVT data—without which predictions are based on a group-average model (Ramakrishnan, Wesensten, Kamimori, et al., 2016)—are required to individualize the alertness prediction model.

The *2B-Alert* App generates individualized and group-average alertness predictions for mean response time (RT) as the PVT statistic. We chose mean RT because it is one of the most frequently used PVT statistics (Basner & Dinges, 2011), and because it allows for the most accurate estimation of individual-specific predictions of alertness under both TSD and CSR (Liu et al., 2017; Ramakrishnan et al., 2015). Nevertheless, the app also stores raw RT data. Figure 1e,f show the screens accessed by the “Status” and “Prediction” buttons, respectively, in the main menu, which provide individualized or group-average alertness predictions for mean RT as a function of time of day.

### 2.3 | Initial conditions and model individualization

The predictions of alertness by the UMP are based on: (a) sleep/wake history; (b) caffeine dosage, absorption rate and time of consumption; and (c) time of day (Ramakrishnan, Wesensten, Balkin, et al., 2016; Ramakrishnan, Wesensten, Kamimori, et al., 2016). Absent of prior data, the predictions assume that the user slept 8 hr (23:00–07:00 hr) on the previous night and has no accumulated sleep debt. In addition, it uses initial model parameters tuned to provide group-average predictions based on 10-min PVTs (Ramakrishnan, Wesensten, Kamimori, et al., 2016; Reifman et al., 2016). Thereafter, the model predictions increasingly reflect the inputs (sleep and caffeine) entered by the user, but continue to reflect group-average alertness until the user performs a PVT.

Immediately after the first PVT session, the app automatically adapts the UMP parameters on the fly, using a Bayesian learning algorithm (Liu et al., 2017). At this point, the predictions start to reflect the user’s trait-like response to sleep deprivation. Subsequently, after each intermittent PVT, the app progressively adapts the model parameters to match the PVT results. In the process, it “learns” the user’s response to different sleep and caffeine schedules as a function of time of day.

## 2.4 | Supported mobile computing devices

The *2B-Alert* App was implemented in two operating systems, iOS and Android, supporting iPhone 6s and iPad Air 2 running iOS 9.3+ and Samsung Note 4 running Android 6.0.1+, respectively. Both implementations provide the same features and a consistent user experience. Supporting Information Data S1, Section II, provides a detailed description of the software architecture.

## 2.5 | PVT settings and latency characterization

Users perform PVTs via the “PVT Session” interface (Figure 1a,d), where “Settings” allows for the selection of a PVT session duration of 3, 5 or 10 min and customization of the inter-stimulus interval (ISI), with a minimum delay of 1 or 2 s and a maximum delay ranging from 4 to 10 s. Using the customized minimum and maximum delays, the app randomizes the ISI for each stimulus presentation. The reported mean RT accounts for (i.e. it subtracts) the mobile device and software latency, which we characterized using the same RTBox apparatus (Li, et al., 2010) used to develop the PC-PVT and PC-PVT 2.0 software (Khitrov et al., 2014; Reifman, et al., 2018) The

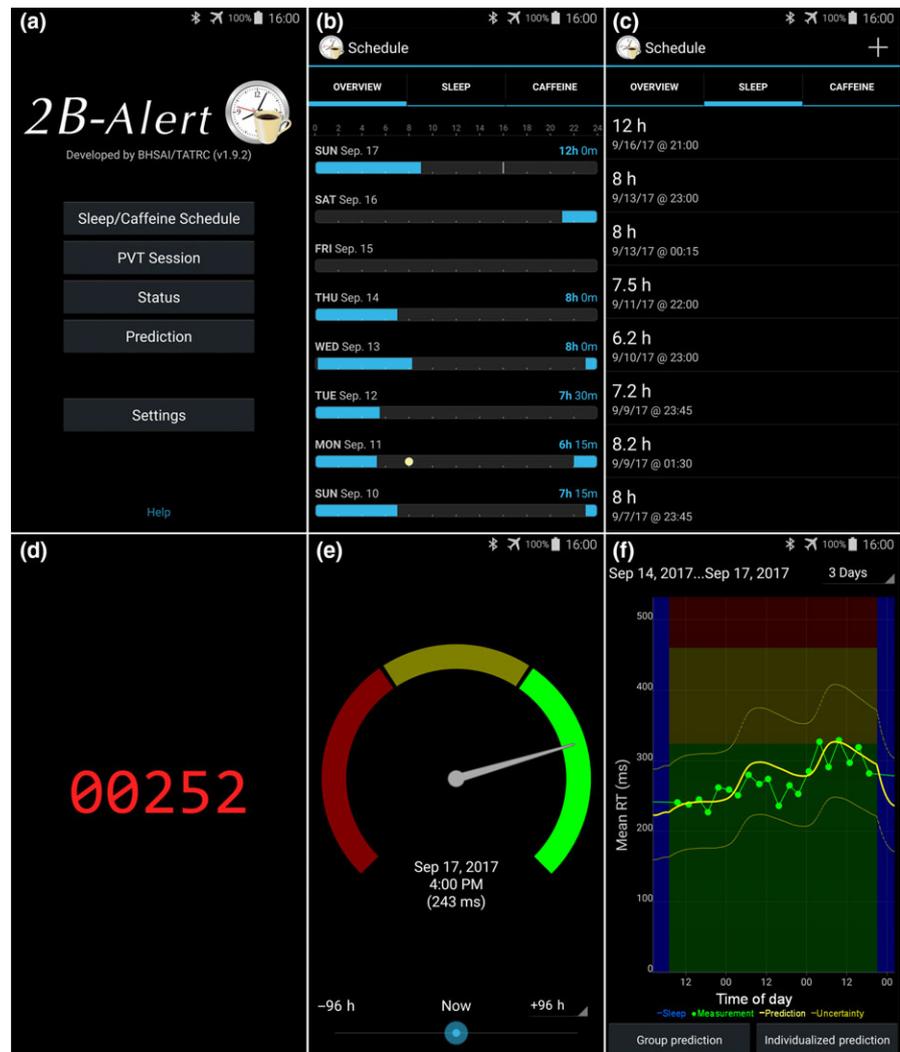
hardware–software RT latency for the *2B-Alert* App was 58 and 68 ms for the iPhone 6s and iPad Air 2, respectively, running iOS 10.3.2, and 79 ms for the Samsung Note 4 running Android 6.0.1. We also assessed whether RT latencies depend on the orientation of the smartphone (portrait vs. landscape) and the finger used for responding (index vs. thumb; Arsintescu, Mulligan, & Flynn-Evans, 2017), but found no significant statistical differences among the configurations. Nonetheless, we recommend that PVTs in the app be performed consistently, always using the same configuration.

## 2.6 | Validation study

To assess the PVT and real-time individualized predictions of alertness in the *2B-Alert* App, we performed a prospective sleep-loss study, where we collected and stored all test and prediction data within the app.

We recruited 21 healthy participants (14 men and seven women; 18 civilians and three active-duty personnel), ranging in age from 18 to 34 years [mean = 24.6 years and standard deviation (*SD*) = 4.6 years], who were screened for sleep disorders and mental health problems. The study was approved by the Walter Reed Army

**FIGURE 1** Screen captures of the main graphical user interfaces of the *2B-Alert* App, including the main menu (a), overview of sleep (light blue horizontal bars) and caffeine schedules (yellow dot) accessed via the “Sleep/Caffeine Schedule” button on the main menu (b), sleep input (showing the duration and start of each sleep episode) accessed via the “Sleep” link on the “Schedule” page (c), psychomotor vigilance test (PVT) stimulus accessed via the “PVT Session” button on the main menu (d), individualized (or group-average) prediction of mean response time (RT) PVT performance through a dial and needle representation accessed via the “Status” button on the main menu (e), and individualized (or group-average) prediction of mean RT PVT performance as a function of time of day accessed via the “Prediction” button on the main menu (f)



Institute of Research Human Use Committee (Silver Spring, MD, USA). Written informed consent was obtained from each subject prior to their participation.

During each of the 13 days prior to the in-laboratory phase, participants completed a sleep/wake and caffeine diary at home using the app in a Samsung Note 4, and continually wore a Philips Respironics Actiwatch-2 wrist actigraphy watch. During this period, participants slept an average of 7.7 hr per night ( $SD = 1.7$  hr), as estimated by actigraphy. Participants were allowed to consume no more than 400 mg of caffeine per day. Each day, they used the app to perform five to six 5-min PVTs (ISI: 1–4 s), once every 2–3 hr.

On the 13th day, participants reported to the laboratory at 19:00 hr, went to bed at 23:00 hr, woke up at 07:00 hr the next morning, and then underwent 62 hr of TSD followed by 1 night of recovery sleep (12 hr in bed). Participants manually entered the corresponding hours of sleep into the app (Figure 1b,c). During this laboratory phase, participants performed three different versions of the 5-min PVT: PC-PVT and Gamified PC-PVT (not reported here) using a standalone PC, and 2B-Alert PVT (each once every 3 hr, order counterbalanced, with a 5-min break between versions). In all, each participant performed 23 tests per version (20 during TSD, three after recovery). All data (raw PVT and predictions) were stored in the PC and app, and retrieved after the study was completed.

## 2.7 | Statistical analyses

To assess the reliability of the PVT statistics obtained with the app, we compared them against those collected by the PC-PVT (Khitrov et al., 2014) for each of seven PVT statistics [mean RT, median RT, slowest 10% RT, speed, lapse 500 ms (number of RTs > 500 ms), lapse 355 ms (number of RTs > 355 ms), and number of false starts]. We computed: (a) paired differences for three levels of sleep deprivation; (b) Pearson's correlation using data from all 23 PVTs during the in-laboratory phase; (c) Spearman's rank-order correlation of data averaged over the 20 tests during TSD; and (d) effect size (Cohen's  $d$ ) between group PVT performance at baseline (0–16 hr of TSD) and that after extended sleep loss (40–62 hr of TSD; McGrath & Meyer, 2006).

To assess the app's ability to learn and accurately predict an individual's response to sleep loss in real time, we calculated the root mean squared errors (RMSEs) between the app-measured mean RTs and the real-time app-predicted mean RTs. To assess how well the 2B-Alert App progressively learned each individual's trait-like response to sleep loss, we compared the RMSEs over the last 24 hr of TSD as a function of the number of PVTs used for model individualization. We also computed the "best-fit" model estimate for each individual, which we obtained retrospectively by fitting the UMP while using all available 2B-Alert PVT data for an individual.

## 2.8 | Access

The 2B-Alert App is available for collaborative research on a case-by-case basis.

## 3 | RESULTS

### 3.1 | User interface and functionalities

The main menu of the 2B-Alert App provides access to its various functionalities (Figure 1a). The "Sleep/Caffeine Schedule" button provides for the visualization of the user's sleep/wake and caffeine consumption schedule history ("overview" in Figure 1b), and separate interfaces for inputting and editing both "sleep" (Figure 1c) and "caffeine" schedules (not shown). Figure 1b shows the 8-day schedule for subject #11, corresponding to the last 4 nights of the at-home phase (10–13 September), the 8 hr of time in bed during the first night in the laboratory, the 62 hr of TSD, followed by the 12 hr of time in bed during the recovery night.

The "Status" button in Figure 1a leads to one of the two interface screens for visualizing alertness predictions. The status interface in Figure 1e shows a horseshoe-shaped dial, with the needle indicating mean RT performance (243 ms) on the corresponding date (17 September 2017) and time of day (16:00 hr) associated with the round blue time marker at the bottom of the screen. By horizontally moving the marker, users can access predictions up to 96 hr into the future, or the past. The dial has three colours (green, yellow and red), indicating increasing levels of alertness vulnerability going from green to red, with the green-to-yellow threshold being equivalent to a 0.06% blood alcohol concentration (BAC), whereas the yellow-to-red threshold corresponds to a BAC of 0.08% (Dawson & Reid, 1997; Reifman et al., 2016; Williamson, et al., 2001).

The "Prediction" button in Figure 1a provides access to the second interface screen for visualizing alertness predictions. Figure 1f shows the resulting screen, which depicts the predicted mean RT (on the y-axis) as a function of the time of day (on the x-axis). At the bottom of the screen, the user can choose to visualize "Group Prediction" or "Individualized Prediction" (the one selected here). The yellow line indicates UMP-predicted alertness and the dashed yellow lines denote the associated 95% prediction intervals, whereas the green dots indicate PVT mean RT data used to individualize the model parameters and the dark blue vertical bars indicate sleep periods. The figure shows PVT data and predictions for the 62 hr of TSD for subject #11 (Figure 1b).

### 3.2 | Validation results

#### 3.2.1 | 2B-Alert PVT measurements

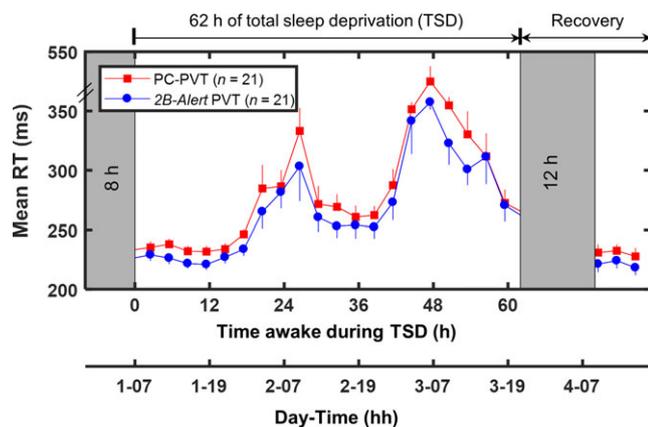
To validate the 2B-Alert PVT against the PC-PVT, we quantified its bias, computed their correlations in value and rank, and compared effect sizes of sleep loss between the two devices on 5-min tests, for each of seven PVT statistics. Figure 2 shows the temporal profile of mean RT for the two devices, while Table 1 shows the means and SDs of paired differences across three different TSD periods, Pearson's and Spearman's correlation coefficients, and effect sizes of sleep loss. The 2B-Alert PVT generally resulted in lower impairment scores than the PC-PVT, yielding statistically significant differences

for mean RT, median RT and speed for each of three TSD periods ( $p < 0.05$ , Wilcoxon signed rank test). Each statistic was well correlated ( $p < 0.001$ ) in value and rank across versions, with Pearson's coefficient exceeding 0.5 and Spearman's coefficient exceeding 0.75 for all statistics. Lastly, except for false starts, the effect sizes of sleep loss on 2B-Alert PVT-derived statistics (1.53–1.93) were comparable to those on PC-PVT-derived statistics (1.49–2.20). Importantly, the average psychomotor vigilance after extended sleep loss (40–62 hr) was significantly worse than that at baseline (0–16 hr) for all statistics in both versions ( $p < 0.01$ , Wilcoxon signed rank test; results not shown in Table 1).

### 3.2.2 | 2B-Alert PVT predictions

To validate the ability of the app to learn and predict an individual's response to sleep loss in real time, we analysed the predictions of mean RT over the last 24 hr of TSD as a function of the number of PVTs used for model individualization. Figure 3 shows the 2B-Alert PVT data and best-fit model estimate for the last day of the at-home phase of the study, the 62 hr of TSD, and the recovery day for three subjects (#7, #9 and #11) representing three broadly different trait-like responses to sleep loss. The retrospectively obtained best-fit model estimate indicates the upper-limit accuracy we can expect from the real-time app predictions. The figure also shows individualized predictions after the at-home phase, as well as after 8 and 12 PVTs during TSD. For each subject, as the number of available measurements increased, the predictions progressively converged to those of the best-fit estimate.

Figure 4 shows the average RMSE of the individualized predictions over the 21 subjects as a function of the number of PVT measurements available—and used by the app—during TSD for model customization (Supporting information Table S1 provides RMSEs for each subject). The accuracy of the predictions gradually increased with the number of PVT measurements used for model



**FIGURE 2** Group-averaged ( $n = 21$ ) psychomotor vigilance test (PVT) mean response time (RT) data (standard errors) from the PC-PVT (solid red squares) and the 2B-Alert PVT (solid blue circles) during 62 hr of total sleep deprivation (TSD) and following 12 hr of recovery sleep

customization, with the average RMSE falling within ~10 ms of the best-fit RMSE after 12 PVTs during TSD.

## 4 | DISCUSSION

The 2B-Alert App is the first mobile application that progressively *learns* and with increasing accuracy *predicts* in real time the effects of sleep/wake schedules, caffeine consumption and time of day on an individual's alertness. Here we described its implementation in two smartphone platforms, and validated its reliability in measuring and predicting PVT performance during a prospective 62-hr TSD study.

The temporal profiles of the 2B-Alert PVT and the PC-PVT showed that TSD induced alertness impairment, increasing mean RTs and degrading performance on other PVT statistics. We found that both devices captured circadian effects and yielded well-correlated 5-min PVT statistics (Figure 2; Table 1), detecting both within- and between-subject variations in the temporal profiles of performance across 62 hr of TSD. Compared with the PC-PVT, the 2B-Alert PVT yielded faster RTs and more false starts, presumably due to device-dependent factors (Honn, Riedy, & Grant, 2015). Grant and colleagues observed similar results, with 3-min PVTs on a smartphone app yielding faster RTs, fewer lapses and more false starts than on a tablet (Grant et al., 2017). Except for false starts, effect sizes of 2B-Alert PVT-derived statistics were generally slightly lower or similar to those derived from PC-PVT, a trend that was comparable to that observed in Brunet et al. (2017). Interestingly, upon reducing the lapse threshold from 500 to 355 ms, the effect size of the lapse statistic decreased from 1.87 to 1.66 for the PC-PVT, while increasing marginally from 1.91 to 1.93 for the 2B-Alert PVT, suggesting that the app is equally sensitive to sleep loss regardless of lapse threshold. Most importantly, sleep loss had a large effect (Cohen's  $d > 1.00$ ) on all 2B-Alert PVT-derived statistics. Together, our results support the use of the app as a viable tool for measuring alertness impairment due to sleep loss.

The 2B-Alert App progressively and automatically learned in real time each individual's trait-like response to sleep loss during TSD, yielding increasingly more accurate predictions of alertness as the number of PVT measurements available for model customization accrued. Reassuringly, the average prediction error at the end of the 62-hr TSD challenge using the app (46 ms) was equivalent to the error obtained in a post hoc analysis of a previous 64-hr TSD study (45 ms; Liu et al., 2017). Of note, here we presented results computed in real time, using all collected data and all subjects, just as such an application is intended to work in the real world. This is in stark contrast to all previously reported model predictions of alertness by us and others, where computations are performed off line upon completion of a study, often after eliminating outlier data and anomalous subjects.

Perhaps unsurprisingly, the app could not learn an individual's response to sleep loss during the at-home phase of the study, when the habitual sleep/wake schedule (which averaged almost 8 hr of sleep per night) was maintained (Figure 4). It took at least 12 PVT

**TABLE 1** Comparison of 2B-Alert PVT against PC-PVT for seven PVT statistics

PVT statistic	Mean paired difference (SD) (PC-PVT—2B-Alert PVT)			Correlation coefficient		Effect size (Cohen's <i>d</i> )	
	0–16 hr ( <i>n</i> = 105)	16–40 hr ( <i>n</i> = 168)	40–62 hr ( <i>n</i> = 147)	Pearson ( <i>n</i> = 483)	Spearman ( <i>n</i> = 21)	PC-PVT	2B-Alert PVT
Mean RT (ms)	9 (21) <sup>a</sup>	14 (56) <sup>a</sup>	25 (104) <sup>a</sup>	0.72	0.86	1.99	1.88
Median RT (ms)	8 (18) <sup>a</sup>	12 (28) <sup>a</sup>	18 (60) <sup>a</sup>	0.61	0.83	1.49	1.53
Slowest 10% RT (ms)	8 (75) <sup>a</sup>	18 (281) <sup>a</sup>	55 (499)	0.71	0.77	2.20	1.86
Speed (s <sup>-1</sup> )	-0.2 (0.4) <sup>a</sup>	-0.2 (0.3) <sup>a</sup>	-0.2 (0.5) <sup>a</sup>	0.79	0.84	1.76	1.66
Lapse 500 ms (#)	-0.2 (0.9) <sup>a</sup>	0.2 (3.2)	0.7 (5.3)	0.76	0.78	1.87	1.91
Lapse 355 ms (#)	1.0 (3.6)	3.0 (5.5) <sup>a</sup>	2.8 (7.9) <sup>a</sup>	0.77	0.84	1.66	1.93
False starts (#)	-0.6 (4.3)	-3.6 (7.8) <sup>a</sup>	-6.6 (9.9) <sup>a</sup>	0.56	0.83	0.64	1.19

Note. Shown are the mean (standard deviation [SD]) of paired differences (across three different periods of total sleep deprivation) and correlation coefficients (Pearson and Spearman) between the statistics obtained from the two PVT versions. Also shown are the effect sizes of sleep loss in the two PVT versions. All results are based on 5-min PVTs.

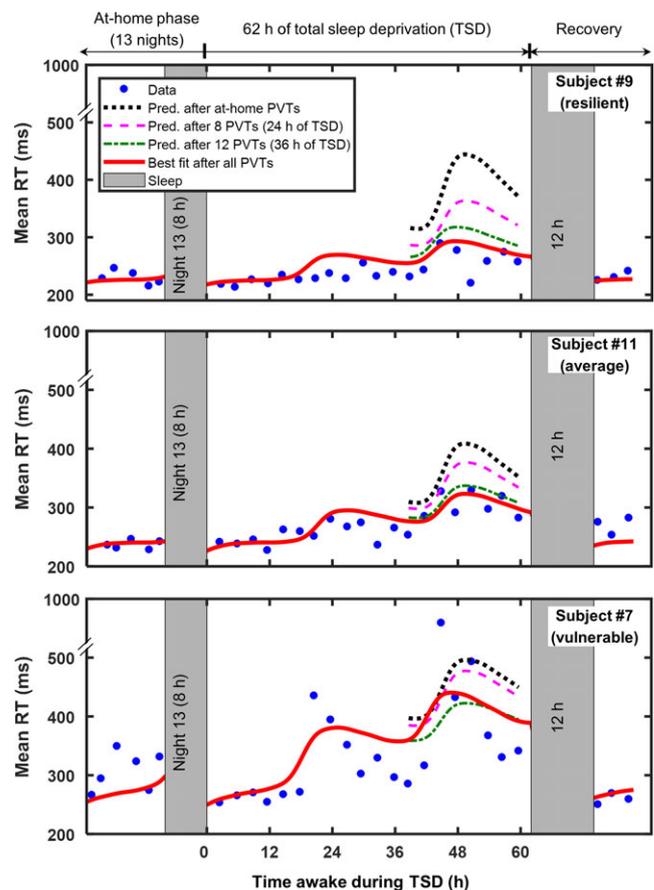
PC: personal computer; PVT: psychomotor vigilance test; RT: response time.

<sup>a</sup>*p* < 0.05, one-tailed Wilcoxon signed rank test.

measurements over the first 36 hr of TSD to generate predictions comparable to those of the best-fit model. Conceivably, the model could have learned faster, using fewer measurements, had the original model been developed using 5-min PVTs, as in this study (as described in Section 2.3, the original model was developed using 10-min PVT data). This is because while the app continually individualizes the model parameters, model individualization is faster, requiring fewer measurements, if the original model parameters are similar to those of the final, individualized model. Nevertheless, in this regard, we also expect the app to be equally capable of predicting alertness impairment using 3-min PVT data. Overall, these results suggest that the app needs PVT measurements during sleep deprivation to learn and accurately predict an individual's trait-like response to limited sleep.

The 2B-Alert App takes as input either actual sleep time or time in bed. The latter might result in the underestimation of alertness impairment, especially when sleep efficiency is low because of exogenous factors, such as ambient lighting and environmental noise, or endogenous factors, such as unfavourable circadian timing. However, because the app individualizes the model parameters using alertness feedback from the user, i.e. through the response to PVTs, we expect that differences between time in bed and actual sleep times will be progressively compensated for as PVTs are conducted. In addition, we performed sensitivity analysis of the prediction model to assess how sleep-time errors affected alertness–impairment predictions, and found that consecutive daily over- or under-estimation of sleep by as much as 30 min per day led to a maximum alertness prediction error of no more than 10%. This suggests that the app is robust to small discrepancies in sleep-recording times.

In contrast to existing personal sleep-tracking devices, the 2B-Alert App can provide *actionable* information to help mitigate the detrimental effects of sleep deprivation on alertness. For example, 40% of US Service members sleep less than 5 hr each night (Luxton et al., 2011; Mysliwiec et al., 2013), leading to substantial chronic sleep debt. In addition, certain military occupations only provide



**FIGURE 3** Mean response time (RT) data for the last day of the at-home phase of the study, 62 hr of total sleep deprivation (TSD), and recovery day for three subjects (#7, #9 and #11), who exhibited different patterns of response to sleep loss. The three dotted and/or dashed lines indicate 2B-Alert App real-time individualized predictions for the last 24 hr of TSD (hours 38–62), using models customized based on psychomotor vigilance test (PVT) data collected during the at-home phase, and after 8 and 12 PVTs during TSD. The best-fit model estimate is shown for comparison by the solid red line

limited and varying opportunities to sleep, while requiring periods of heightened vigilance during wakefulness, which may vary in duration and time of day. Such daily variations in sleep/wake schedules and requirements for peak alertness at both favourable and unfavourable phases of the circadian make it challenging, if not impossible, for individuals to self-administer countermeasures to achieve peak alertness at the desired times of the day for the desired durations. In such a scenario, an already-individualized *2B-Alert* App would use the Service member's recent history of caffeine consumption and sleep/wake schedule to make individualized predictions of future alertness levels, as well as those regarding the efficacy of future caffeine and nap interventions, to determine the one that optimizes alertness for periods of required heightened vigilance.

The UMP has been validated using data from more than a dozen studies, including those investigating sleep loss (Ramakrishnan, Wesensten, Balkin, et al., 2016) and the effects of caffeine during sleep loss (Ramakrishnan, Wesensten, Kamimori, et al., 2016), in both laboratory and field environments. Nevertheless, the model has limitations. The UMP was developed using simple reaction time data (i.e. from a PVT), and the extent to which its predictions generalize to other aspects of neurobehavioural performance is unknown. Also, the validation studies invariably involved a homogenous population of young (<40 years old), healthy adults, and the extent to which the results can be extrapolated to an older, heterogeneous population remains to be determined. Another question concerns the use of the app under conditions of both sleep loss and circadian misalignment. Although we have previously shown that the model accurately predicts the effects of daytime sleep under sleep

deprivation for a group of individuals [e.g. 2-hr daytime naps (Ramakrishnan, Wesensten, Balkin, et al., 2016) and 4 hr of daytime sleep per 24-hr period (Ramakrishnan, Wesensten, Kamimori, et al., 2016)], it is not yet known whether this holds for individual-level predictions. However, because the app continually updates the model parameters and two of the three most sensitive model parameters are used to characterize the circadian component [i.e. the circadian amplitude and phase (Liu et al., 2017)], we expect the app to progressively capture circadian misalignments due to shift work and transmeridian travel. Finally, while part of this study involved an at-home phase, the sleep-deprivation component was conducted in a laboratory. Hence, we need to assess how the app would work in a "real-world" situation, where laboratory-controlled conditions are not applicable.

Requiring a period of sleep deprivation to learn an individual's sleep-loss phenotype could preclude the wide-spread use of the app as a fatigue-management tool. However, TSD, as used in this study, is not required. We have recently shown that the model can learn, in less than a week, the phenotype for a group of individuals who had been challenged with 3 hr of sleep per night (Liu et al., 2017). What would be valuable to know is whether the model can learn the sleep-loss phenotype of an individual under milder sleep-deprivation requirements, such as one week of 5 hr of sleep per night—a condition more commonly experienced by the general population. An affirmative answer to this question would mitigate the limitation mentioned above.

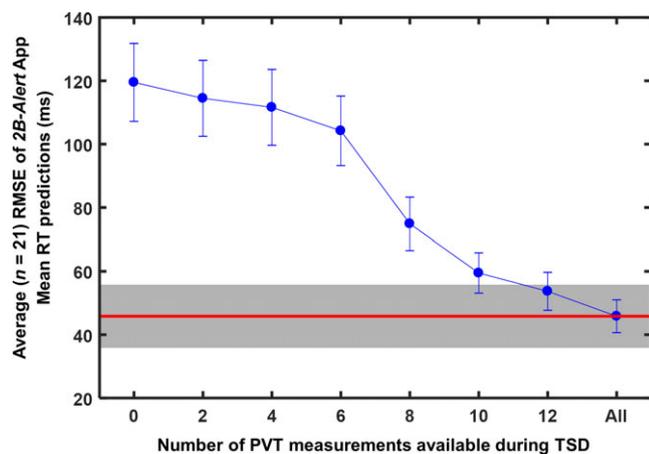
The *2B-Alert* App incorporates the latest scientific findings on sleep restriction, sleep extension, caffeine and recovery sleep, automatically learning an individual's trait-like response to sleep deprivation on the fly, to generate real-time individualized predictions of alertness. Importantly, this allows users to prospectively compare and contrast the efficacy of different nap and caffeine interventions to tailor personalized fatigue management strategies, facilitating the self-management of alertness and safety in both operational and non-operational settings.

## ACKNOWLEDGEMENTS

The authors express their gratitude to Joy Hoffman, Samantha Hornby, Ingrid Lim, Nick Tountas and Nancy Wesensten for providing user feedback throughout the development of the *2B-Alert* App. This work was sponsored by the Military Operational Medicine Research Program of the US Army Medical Research and Materiel Command, Ft. Detrick, MD.

## CONFLICT OF INTEREST

This was not an industry-supported study. The authors have indicated no financial 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 US Army or of the US Department of Defense. This paper has been approved for public release with unlimited distribution.



**FIGURE 4** Average ( $n = 21$ ) root mean squared error (RMSE) of *2B-Alert* App individualized mean response time (RT) predictions for the last 24 hr of total sleep deprivation (TSD), as a function of the number of PVT measurements available (and used by the app) during TSD for model customization. The first (leftmost) RMSE corresponds to predictions based on models customized using the at-home phase of the study. Error bars indicate one standard error. The red horizontal line indicates the average RMSE of the best-fit model, and the grey shaded bar represents a range of  $\pm 10$  ms within the average best-fit RMSE

## AUTHOR CONTRIBUTIONS

JR conceived the research and wrote the paper. SR, JL, AK, KK and MYK developed the algorithms and performed the computations. KK and MYK developed the apps. JR, SR and TJB provided input for system development. TJD and TJB performed the validation studies.

## ORCID

Jaques Reifman  <http://orcid.org/0000-0001-7292-2029>

## REFERENCES

- Arsintescu, L., Mulligan, J. B., & Flynn-Evans, E. E. (2017). Evaluation of a psychomotor vigilance task for touch screen devices. *Human Factors*, *59*, 661–670. <https://doi.org/10.1177/0018720816688394>
- Basner, M., & Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep*, *34*, 581–591. <https://doi.org/10.1093/sleep/34.5.581>
- Brunet, J. F., Dagenais, D., Therrien, M., Gartenberg, D., & Forest, G. (2017). Validation of sleep-2-Peak: A smartphone application that can detect fatigue-related changes in reaction times during sleep deprivation. *Behavior Research Methods*, *49*, 1460–1469. <https://doi.org/10.3758/s13428-016-0802-5>
- Dawson, D., & Reid, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, *388*, 235. <https://doi.org/10.1038/40775>
- Ferguson, T., Rowlands, A. V., Olds, T., & Maher, C. (2015). The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: A cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, *12*, 42. <https://doi.org/10.1186/s12966-015-0201-9>
- Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., & Van Dongen, H. P. A. (2017). 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. *Behavior Research Methods*, *49*, 1020–1029. <https://doi.org/10.3758/s13428-016-0763-8>
- Honn, K. A., Riedy, S. M., & Grant, D. A. (2015). Validation of a portable, touch-screen psychomotor vigilance test. *Aerospace Medicine and Human Performance*, *86*, 428–434. <https://doi.org/10.3357/AMHP.4165.2015>
- Khitrov, M. Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N. J., & Reifman, J. (2014). PC-PVT: A platform for psychomotor vigilance task testing, analysis, and prediction. *Behavior Research Methods*, *46*, 140–147. <https://doi.org/10.3758/s13428-013-0339-9>
- Li, X., Liang, Z., Kleiner, M., & Lu, Z. L. (2010). RTbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*, 212–225. <https://doi.org/10.3758/BRM.42.1.212>
- Liu, J., Ramakrishnan, S., Laxminarayan, S., Balkin, T. J., & Reifman, J. (2017). Real-time individualization of the unified model of performance. *Journal of Sleep Research*, *26*, 820–831. <https://doi.org/10.1111/jsr.12535>
- Luxton, D. D., Greenburg, D., Ryan, J., Niven, A., Wheeler, G., & Mysliwiec, V. (2011). Prevalence and impact of short sleep duration in redeployed OIF soldiers. *Sleep*, *34*, 1189–1195. <https://doi.org/10.5665/SLEEP.1236>
- McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of  $r$  and  $d$ . *Psychological Methods*, *11*, 386–401. <https://doi.org/10.1037/1082-989X.11.4.386>
- Mysliwiec, V., McGraw, L., Pierce, R., Smith, P., Trapp, B., & Roth, B. J. (2013). Sleep disorders and associated medical comorbidities in active duty military personnel. *Sleep*, *36*, 167–174. <https://doi.org/10.5665/sleep.2364>
- Ramakrishnan, S., Laxminarayan, S., Wesensten, N. J., Kamimori, G. H., Balkin, T. J., & Reifman, J. (2014). Dose-dependent model of caffeine effects on human vigilance during total sleep deprivation. *Journal of Theoretical Biology*, *358*, 11–24. <https://doi.org/10.1016/j.jtbi.2014.05.017>
- Ramakrishnan, S., Lu, W., Laxminarayan, S., Wesensten, N. J., Rupp, T. L., Balkin, T. J., & Reifman, J. (2015). Can a mathematical model predict an individual's trait-like response to both total and partial sleep loss? *Journal of Sleep Research*, *24*, 262–269. <https://doi.org/10.1111/jsr.12272>
- Ramakrishnan, S., Wesensten, N. J., Balkin, T. J., & Reifman, J. (2016). A unified model of performance: Validation of its predictions across different sleep/wake schedules. *Sleep*, *39*, 249–262.
- Ramakrishnan, S., Wesensten, N. J., Kamimori, G. H., Moon, J. E., Balkin, T. J., & Reifman, J. (2016). A unified model of performance for predicting the effects of sleep and caffeine. *Sleep*, *39*, 1827–1841.
- Reifman, J., Kumar, K., Khitrov, M. Y., Liu, J., & Ramakrishnan, S. (2018). PC-PVT 2.0: An updated platform for psychomotor vigilance task testing, analysis, prediction, and visualization. *Journal of Neuroscience Methods*, *304*, 39–45. <https://doi.org/10.1016/j.jneumeth.2018.04.007>
- Reifman, J., Kumar, K., Wesensten, N. J., Tountas, N. A., Balkin, T. J., & Ramakrishnan, S. (2016). 2B-Alert Web: An open-access tool for predicting the effects of sleep/wake schedules and caffeine consumption on neurobehavioral performance. *Sleep*, *39*, 2157–2159. <https://doi.org/10.5665/sleep.6318>
- Williamson, A. M., Feyer, A. M., Mattick, R. P., Friswell, R., & Finlay-Brown, S. (2001). Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. *Accident Analysis & Prevention*, *33*, 313–326. [https://doi.org/10.1016/S0001-4575\(00\)00045-2](https://doi.org/10.1016/S0001-4575(00)00045-2)

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Reifman J, Ramakrishnan S, Liu J, et al. 2B-Alert App: A mobile application for real-time individualized prediction of alertness. *J Sleep Res*. 2018; e12725. <https://doi.org/10.1111/jsr.12725>