



ORIGINAL ARTICLE

Response speed measurements on the psychomotor vigilance test: how precise is precise enough?

Mathias Basner^{1,*†,◊}, Tyler M. Moore^{2,†}, Jad Nasrini¹, Ruben C. Gur² and David F. Dinges¹

¹Unit of Experimental Psychiatry, Division of Sleep and Chronobiology, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA and ²Brain Behavior Laboratory, Neuropsychiatry Section, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[†]These authors contributed equally.

*Corresponding author. Mathias Basner, Unit of Experimental Psychiatry, Division of Sleep and Chronobiology, Perelman School of Medicine, University of Pennsylvania, 1019 Blockley Hall, 423 Guardian Drive, Philadelphia, PA. Email: basner@penmedicine.upenn.edu.

Abstract

Study Objectives: The psychomotor vigilance test (PVT) is frequently used to measure behavioral alertness in sleep research on various software and hardware platforms. In contrast to many other cognitive tests, PVT response time (RT) shifts of a few milliseconds can be meaningful. It is, therefore, important to use calibrated systems, but calibration standards are currently missing. This study investigated the influence of system latency bias and its variability on two frequently used PVT performance metrics, attentional lapses (RTs ≥ 500 ms) and response speed, in sleep-deprived and alert participants.

Methods: PVT data from one acute total ($N = 31$ participants) and one chronic partial ($N = 43$ participants) sleep deprivation protocol were the basis for simulations in which response bias (± 15 ms) and its variability (0–50 ms) were systematically varied and transgressions of predefined thresholds (i.e. ± 1 for lapses, $\pm 0.1/s$ for response speed) recorded.

Results: Both increasing bias and its variability caused deviations from true scores that were higher for the number of lapses in sleep-deprived participants and for response speed in alert participants. Threshold transgressions were typically rare (i.e. $< 5\%$) if system latency bias was less than ± 5 ms and its standard deviation was ≤ 10 ms.

Conclusions: A bias of ± 5 ms with a standard deviation of ≤ 10 ms could be considered maximally allowable margins for calibrating PVT systems for timing accuracy. Future studies should report the average system latency and its standard deviation in addition to adhering to published standards for administering and analyzing the PVT.

Statement of Significance

The psychomotor vigilance test (PVT) is a sensitive and popular assay of the behavioral effects of sleep restriction and circadian misalignment. However, standardization of the PVT is poor both in terms of how to administer and analyze the PVT, as well as in terms of accuracy requirements for the chosen software and hardware platforms. To our knowledge, this is the first systematic investigation on the effects of a biased system latency and its variability on frequently used PVT outcome metrics in alert and sleep-deprived participants. It demonstrates that system latency bias should not exceed ± 5 ms and its variability should not exceed 10 ms to guarantee valid inferences from PVT trials.

Key words: PVT; sleep restriction; sleep deprivation; vigilance; attention; calibration

Submitted: 21 February, 2020; Revised: 18 May, 2020

© Sleep Research Society 2020. Published by Oxford University Press on behalf of the Sleep Research Society. All rights reserved. For permissions, please e-mail journals.permissions@oup.com.

Introduction

The psychomotor vigilance test (PVT) [1] is a widely used measure of behavioral alertness, owing in large part to the combination of its high sensitivity to sleep deprivation [2, 3] and negligible aptitude and practice effects [4]. The standard 10-minute PVT measures sustained or vigilant attention by recording response times (RTs) to visual (or auditory) stimuli that occur at random inter-stimulus intervals (ISI, 2–10 s in the standard 10-min PVT, including a 1-s feedback period during which the RT to the last stimulus is displayed) [1, 5]. RT to stimuli attended to has been used since the late 19th century in sleep deprivation research [6, 7] because it offers a simple way to track changes in behavioral alertness caused by inadequate sleep, without the confounding effects of aptitude and learning [2–4, 8]. Moreover, the 10-minute PVT has been highly reliable, with intra-class correlations for key metrics such as lapses reporting test-retest reliability above 0.8 [3]. PVT performance also has ecological validity in that it can reflect real-world risks, because deficits in sustained attention and timely reactions adversely affect many operational tasks, especially those in which work-paced or timely responses are essential (e.g. stable vigilant attention is critical for safe performance in all transportation modes, many security-related tasks, and a wide range of industrial procedures). Lapses in attention as measured by the PVT can occur when fatigue is caused by either sleep loss or time on task [9, 10], which are the two factors that make up virtually all theoretical models of fatigue in real-world performance. There is a large body of literature on attentional deficits having serious consequences in applied settings [11–14].

Sleep deprivation induces reliable changes in PVT performance, causing an overall slowing of RTs, a steady increase in the number of errors of omission (i.e. lapses of attention, historically defined as RTs \geq twice the mean RT or 500 ms), and a more modest increase in errors of commission (i.e. responses without a stimulus, or responses that reflect false starts [FS]) [15]. These effects can increase as task duration increases (the so-called time-on-task effect or vigilance decrement) [16]. The 10-minute PVT has been sensitive to both acute total sleep deprivation (TSD) [8, 15, 17] and chronic partial sleep deprivation (PSD) [15, 17–21]. It is affected by sleep homeostatic and circadian drives and their interactions [22], revealing large but stable inter-participant variability in the response to sleep loss [23]. The PVT demonstrates effects of jet lag and shift work [24] and improvements in alertness after wake-promoting interventions [25, 26] and recovery from sleep loss [27, 28] and after initiation of continuous positive airway pressure (CPAP) treatment in patients with obstructive sleep apnea [29]. The PVT is often used as a “gold standard” measure for the neurobehavioral effects of sleep loss, against which other biomarkers or fatigue detection technologies are compared [30, 31].

However, despite its performance simplicity, the complexity of developing a valid and reliable PVT is often underestimated. In contrast to most other cognitive tests, RT shifts of a few milliseconds can be meaningful on the PVT. For example, the dataset used for the simulations described below contained 1,835 lapses. If we subtract 5 ms from each RT, we end up with 1,778 lapses (drop of 3.1%). If we add 5 ms, we end up with 1,909 lapses (increase of 4.0%). It is, therefore, important to use calibrated software and hardware. Numerous versions of the PVT are available, commercially or for free, across several hardware

platforms. Researchers also frequently program their own version of the PVT. It is, therefore, unclear how data produced on different versions of the PVT compare across studies. This is especially true for computerized versions of the PVT [32, 33] that use a mouse or a keyboard for response input relative to those that use a touchscreen. For example, both the orientation of a smartphone and the input method (e.g. tapping the screen vs. swiping) have been shown to influence RTs [34, 35].

Further complicating the problem, different PVT platforms and devices may have unique response latencies, that is, the time it takes for a stimulus to be displayed plus the time it takes for a response (e.g. pushing down the spacebar) to be registered by the system. Screens refresh at different rates (e.g. 60 Hz), and if the refresh cycle has just passed the PVT response box when the stimulus is initiated by the software, it may take another full cycle before the stimulus is displayed (well-programmed software can take this refresh cycle into account). Also, as a way to save energy, many smartphones and tablets will reduce the screen polling rate if the user has not engaged the screen for a specified time, which can lead to response binning (e.g. in 16.6 ms bins for a polling rate of 60 Hz). During calibration, the average response latency is typically determined by precisely measuring stimulus onset and user response (e.g. with a high-speed camera), and comparing the thus determined true RT to the one registered by the system. This process is repeated for several stimuli, and the average latency is then subtracted from RTs registered by the system [34].

It is currently unclear how errors in PVT system calibration may affect PVT validity. For example, if the calibration overestimates average response latency by 5 ms, are PVT outcome metrics still valid? Furthermore, the system latency is typically the result of several serial processes that may sample at different frequencies, which introduces variability of system response latencies that may be critical for the validity of the PVT. For example, Arsintescu and colleagues [34] evaluated the response latency of a smartphone version of the PVT with high-speed cameras. They found an average response latency of 68.5 ms with a standard deviation of 18.1 ms.

To our knowledge, no study to date has investigated the effect of a systematic bias in system latency (and latency variability) on the validity of PVT results, and these were the objectives of our study. As PVT result validity may be affected differently in alert and sleep-deprived participants, we used data from both acute total and chronic PSD studies and conditions in which participants were well-rested (see below). We investigated the effects of system latency bias and variability on two standard outcomes of the PVT:

1. Lapses of attention (defined as RTs \geq 500 ms) are errors of omission and frequently used as the primary outcome of the PVT [5].
2. Response speed is the reciprocal transform of RTs and has been shown to have favorable statistical properties and high sensitivity to sleep loss relative to the other standard PVT outcomes [5, 36].

Methods

Participants and protocol

The following descriptions of the TSD and PSD protocols as well as PVT administration are adapted from Basner and Dinges [37].

Acute TSD protocol

TSD data were gathered on $N = 36$ participants in a study on the effects of night work and sleep loss on threat detection performance on a simulated luggage screening task [5]. Study participants stayed in the research lab for five consecutive days that included a 33-hour period of TSD. After inspection of the raw PVT data, four participants (11.1%) were excluded from the analysis due to nonadherence or excessive fatigue during the baseline period (i.e. first 16 h of wakefulness). Another participant withdrew after 26 hours awake. Therefore, $N = 31$ participants (mean age \pm SD = 31.1 ± 7.3 y, 18 female) contributed to the analyses presented here. The study started at 08:00 am on day 1 and ended at 08:00 am on day 5. A 33-hour period of TSD started either on day 2 ($N = 22$) or on day 3 ($N = 9$) of the study. On days preceding or following the TSD period, participants had 8-hour sleep opportunities between 12:00 am and 08:00 am. The first sleep period was monitored polysomnographically to exclude possible sleep disorders.

Chronic PSD protocol

PSD data were obtained from $N = 47$ healthy adults in a laboratory protocol involving 5 consecutive nights of sleep restricted to 4 hours per night (04:00 am to 08:00 am period) following two baseline nights with 10 hours of time in bed each. Three participants (6.4%) were excluded from the analysis due to nonadherence or excessive fatigue during baseline data collection. One additional participant had no valid baseline data. Therefore, $N = 43$ participants (16 females) who averaged 30.5 ± 7.3 years (mean \pm SD) contributed to the analyses presented here. A detailed description of the experimental procedures is published elsewhere [28].

Both protocols

In both TSD and PSD experiments, participants were free of acute and chronic medical and psychological conditions, as established by interviews, clinical history, questionnaires, physical exams, and blood and urine tests. They were studied in small groups (4–5) while they remained in the Sleep and Chronobiology Laboratory at the Hospital of the University of Pennsylvania. Throughout both the experiments, participants were continuously monitored by trained staff to ensure adherence to each experimental protocol. They wore wrist actigraphs throughout each protocol. Meals were provided at regular times throughout the protocol, caffeinated foods and drinks were not allowed, and light levels in the laboratory were held constant during scheduled wakefulness (<50 lux) and sleep periods (<1 lux). Ambient temperature was maintained between 22 and 24°C.

In both TSD and PSD experiments, participants completed 30-minute bouts of a neurobehavioral test battery every 2 hours during scheduled wakefulness. The same battery was administered in both protocols and included the Digit Symbol Substitution Task, a Visual Memory Task, an NBACK task, a Visual Search and Tracking Task, and the 10-minute PVT, in that order. In the TSD experiment, each participant performed 17 PVTs in total (starting at 09:00 am after a sleep opportunity from 12:00 am to 08:00 am with bout #1 and ending at 05:00 pm on the next day after a night without sleep with bout #17). The data of the TSD protocol were complete, and thus $N = 527$ test bouts contributed to the analysis. Consistent with previous publications [5, 37, 38], we only used the test bouts administered at 12:00 pm, 04:00 pm, and 08:00 pm on each of baseline days 1 and 2 and the days after restriction nights 1 to 5 in the PSD experiment. Of the $N = 903$ scheduled test bouts, $N = 23$ (2.5%) were missing,

and thus $N = 880$ test bouts contributed to the analysis. Between neurobehavioral test bouts, participants were permitted to read, watch movies and television, play cardboard games, and interact with laboratory staff to help them stay awake, but no naps/sleep or vigorous activities (e.g. exercise) were allowed.

In both the experiments, all participants were informed about potential risks of the study, and a written informed consent and IRB approval were obtained prior to the start of the study. They were compensated for their participation and monitored at home with actigraphy, sleep–wake diaries, and time-stamped phone records for time to bed and time awake during the week immediately before the PSD study.

PVT assessments

We utilized a precise computer-based version of the 10-minute PVT, which was performed and analyzed according to the standards set forward in the study of Basner and Dinges [5]. Participants were instructed to monitor a red rectangular box on the computer screen and press a physical response button as soon as a yellow stimulus counter appeared on the screen, which stopped the counter and displayed the RT in milliseconds for a 1-second period. The ISI, defined as the period between the last response and the appearance of the next stimulus, varied randomly from 2 to 10 seconds. The participant was instructed to press the PVT response button as soon as each stimulus appeared, in order to keep the RT as low as possible, but not to press the button before the stimulus appeared (which yielded a “FS” warning on the display).

Bias simulations

To investigate how PVT metrics change when bias is introduced, we simulated a wide range of bias severity and variability, and calculated scores before and after bias was introduced. Simulations proceeded in the following steps:

1. PVT lapses and PVT response speed were calculated based on the raw data, which was considered bias-free for our purposes.
2. Bias was added to the raw data by adding random numbers drawn from a normal distribution with a mean of μ (e.g. 5 ms) and standard deviation of σ (e.g. 10 ms) (see below) to each reaction time.
3. Scores from #1 above were recalculated using the newly biased data, and the differences between the biased and unbiased scores were examined. The key outcome of interest was the proportion of scores that were “too biased,” where too biased (termed a “violation”) meant that the absolute difference between the biased and unbiased scores was larger than an a priori specified cutoff value. For lapses, this cutoff value was set to 1, meaning if the biased and unbiased scores differed by more than ± 1 lapse, that score was a violation (i.e. too biased). The violation cutoff for response speed was set to 0.1 s^{-1} . These cutoffs were informed by experience of the authors in analyzing PVT data as well as simulations that estimated the variability in PVT outcomes based on the stochastic nature of PVT responses.¹

¹ RTs of a PVT were randomly drawn and assigned to two bins in a way that each bin contained ~50% of RTs. Lapses and response speed were then calculated, and the absolute difference between bins 1 and 2 recorded. The mean absolute difference based on 500 simulations was 1.6 lapses and 0.13 1/s, respectively.

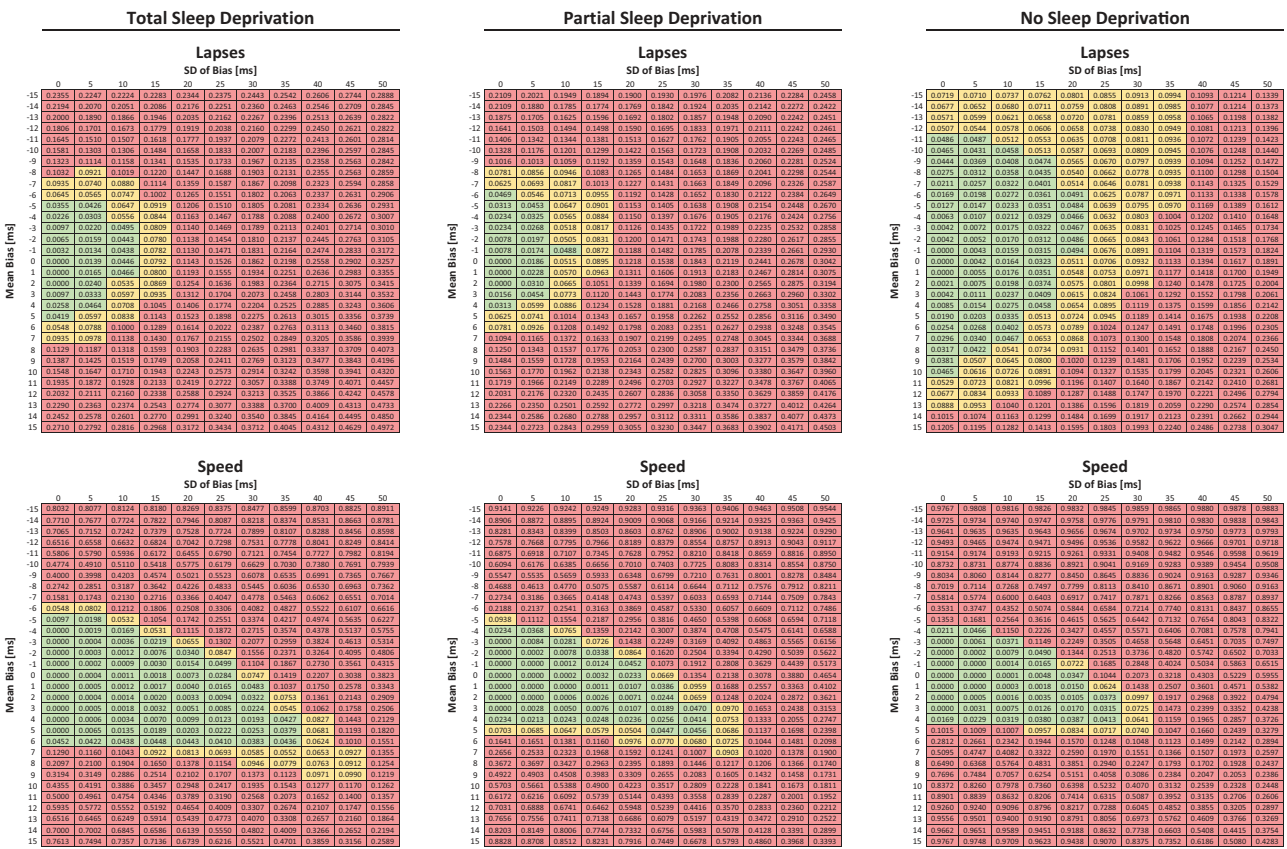


Figure 1. Simulation results for PVT lapses (upper 3 panels) and PVT response speed (lower 3 panels) under different sleep deprivation states: total (left), partial (middle), and no (right) sleep deprivation depending on calibration bias (y-axis) and its standard deviation (x-axis). The numbers in each cell represent the percent of simulations where predefined thresholds of acceptable lapse (1 lapse and 0.1/s, respectively) were exceeded. Cells are color coded based on the % deviation in the following way: green <5%, yellow >5% and <10%, and red $\geq 10\%$.

- Steps #2 and #3 above were repeated 1,000 times, and the mean number was recorded.
- Steps #2, #3, and #4 were repeated after varying μ and σ , specifically, μ ranged from -15 ms to $+15$ ms in 5 ms increments, while σ ranged from 0 to 50 ms in 5 ms increments. All combinations were simulated for a total of $31 \times 11 = 341$ sets of simulations (341,000 simulations total per sample).
- Bland-Altman plots were constructed to examine how the severity of bias changed as overall scores increased. Note that, for calculation of Bland-Altman plots, we used some extreme values of μ not covered in the ranges in #5 above. For example, mean bias as extreme as ± 40 ms was used so as to present a complete range of the phenomena shown in the Bland-Altman plots.
- The above six steps were completed three times, once for each level of sleep deprivation (i.e. TSD, PSD, and no sleep deprivation), for 1.023 million simulations in the full study.

These values shrank to a bias of up to ± 5 ms with a standard deviation of up to 10 ms in the sleep-deprived state. In contrast, the response speed metric was more sensitive to calibration bias and variability in the alert state relative to the sleep-deprived state. However, in both conditions, the maximally tolerated deviation of 0.1 s^{-1} was exceeded in $>5\%$ for a calibration bias outside of ± 5 ms, with more frequent transgressions if the standard deviation of bias was equal to or greater than 20 ms.

Bland-Altman plots are shown for PVT lapses in Figure 2 and for PVT response speed in Figure 3. For both PVT lapses and response speed, the absolute difference between simulated and actual outcomes clearly increased with (1) an increasing number of lapses/higher response speed, (2) increasing calibration bias, and (3) increasing standard deviation of bias. For a bias of up to ± 5 ms and a standard deviation of up to 10 ms, the difference between simulated and actual test scores remained mostly within the predefined limits (i.e. 1 lapse and 0.1 s^{-1}), even for extreme scores (i.e. a high number of lapses and high response speed).

Results

Simulation results are shown for percent violations in Figure 1 and for absolute lapse and response speed values in Supplementary Figure S1. PVT lapses were less affected by calibration bias and precision in the alert compared with the sleep-deprived state. In the alert state, a bias of up to ± 10 ms with a standard deviation of up to 20 ms rarely exceeded the predefined threshold of ± 1 lapse in more than 5% of simulations.

Discussion

This study investigated the influence of the magnitude and variability of measurement error from two frequently used PVT performance outcomes, attentional lapses, and response speed. As expected, both increasing bias and variability of system latency were demonstrated to cause deviations from true scores. Additionally, deviations were higher for test bouts with a high

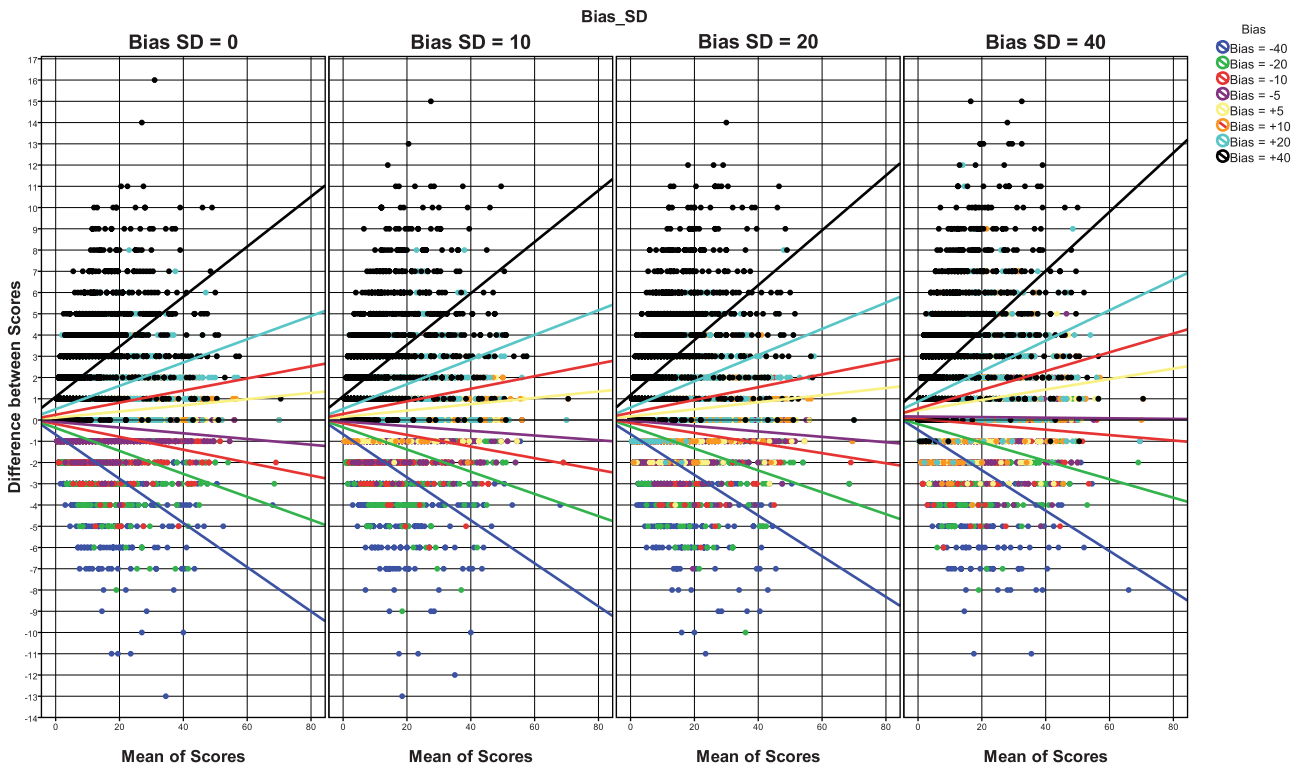


Figure 2. Bland–Altman plots graphing the difference of simulated and actual lapses against the mean of simulated and actual lapses for different response biases (represented by different colors) and standard deviations of response bias (represented in different panels). Each dot reflects one simulation. Lines reflect linear regressions for each bias category.

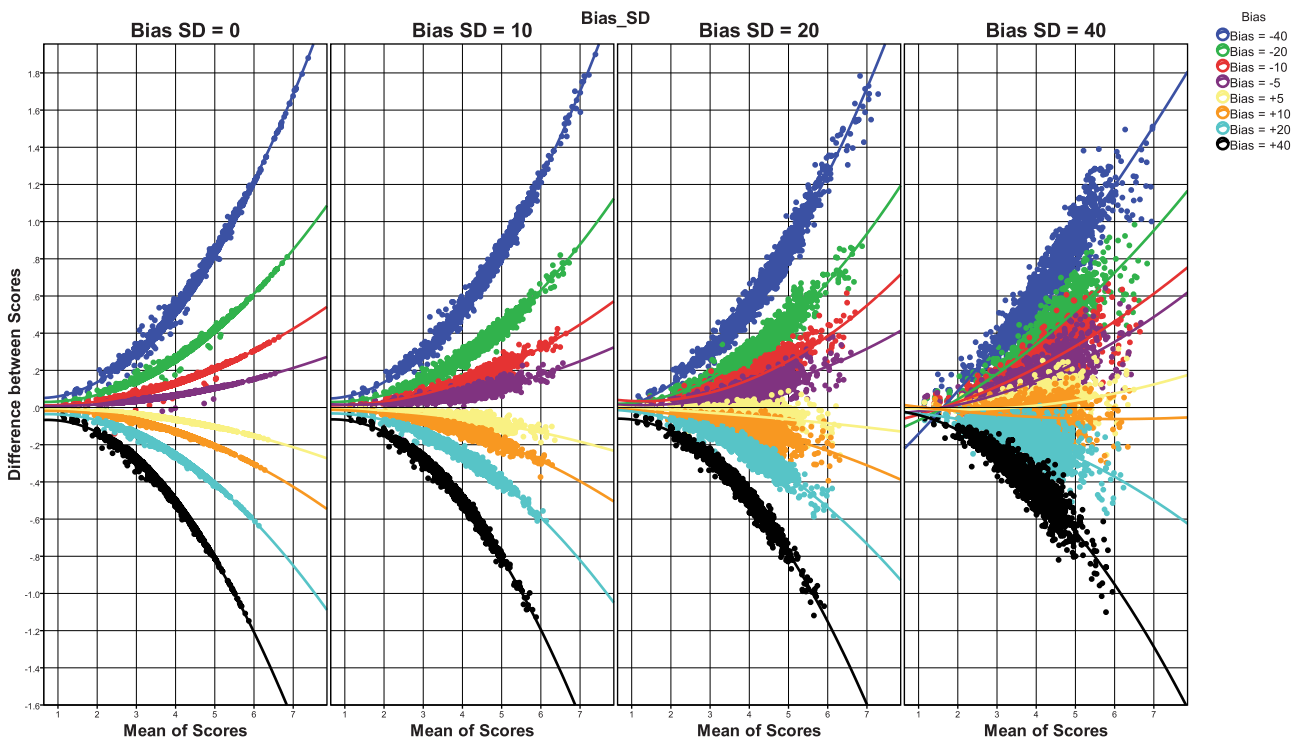


Figure 3. Bland–Altman plots graphing the difference of simulated and actual response speed against the mean of simulated and actual response speed for different response biases (represented by different colors) and standard deviations of response bias (represented in different panels). Each dot reflects one simulation. Lines reflect quadratic fits for each bias category.

number of lapses (i.e. in sleep-deprived participants) or high response speeds (i.e. in alert participants). The latter can be explained by the different properties of the two scores. Lapses on the 10-minute PVT are defined as RTs \geq 500 ms. Alert participants typically average RTs between 200 and 300 ms (i.e. \sim 200–300 ms below the lapse threshold). Even large biases or variable system latency barely affect the number of lapses in this scenario. In contrast, due to the reciprocal transform, the response speed metric shows greater changes in the fast response domain, and larger biases or a variable system latency will, therefore, affect scores of alert participants more than those of sleep-deprived participants.

Transgressions of predefined thresholds (i.e. ± 1 lapse, $\pm 0.1/s$) were typically rare (i.e. $<5\%$) if bias was less than ± 5 ms and the standard deviation of bias was less than 10 ms. Using PVT systems with response latencies that do not meet these criteria will be associated with increasingly biased PVT outcomes depending on the magnitude of the calibration error and system latency variability. These biases may be considered unacceptable depending on the use of the PVT data (e.g. research, fitness-for-duty).

Importantly, correct system calibration also matters for studies in which a participant consistently uses the same hardware. For example, a system that introduces a response latency of 200 ms will, if uncorrected, produce a large number of lapses in the alert state and leave only little room for additional lapses in the sleep-deprived state (i.e. decrease effect size).

Limitations

The investigated participants were healthy, had a restricted age range, and were investigated in a controlled laboratory environment. The results may, therefore, not generalize to non-healthy, older, or younger groups of participants and also may not generalize to operational environments. Our findings may not extend to acute total sleep loss beyond 33 hours, or chronic partial sleep restriction exceeding five consecutive days, or using a sleep restriction paradigm different from 4 hours sleep opportunity per night. Finally, we investigated the effects of response bias and variability only on PVT lapses and response speed, two frequently used outcomes with favorable statistical and interpretational properties. Thus, it is unclear to what extent the findings translate to other PVT outcome metrics.

Conclusions

As expected, simulations based on data derived from alert and sleep-deprived participants demonstrated that both bias and variability of system latency negatively affect the validity of PVT scores. The results of our simulations suggest that a response bias of up to ± 5 ms with a standard deviation of up to ± 10 ms are tolerable for PVT lapses and response speed relative to predefined thresholds. These values could be considered as new standards for the timing accuracy calibration of systems that administer the PVT. Minimally, future studies should report the average system latency and its standard deviation in addition to adhering to published standards for administering and analyzing the PVT [5].

Supplementary Material

Supplementary material is available at SLEEP online.

Funding

This investigation was sponsored by National Aeronautics and Space Administration through NNX16AI53G, the Human Factors Program of the Transportation Security Laboratory, Science and Technology Directorate, U.S. Department of Homeland Security (FAA #04-G-010); by NIH grants R01 NR004281, R01 HL102119, and UL1 RR024134; and in part by the National Space Biomedical Research Institute through NASA NCC 9–58.

Acknowledgments

We thank the participants who participated in the experiments and the faculty and staff who helped acquire the data.

Conflict of interest statement. Financial Disclosure: None; Non-financial disclosure: None.

References

1. Dinges DF, et al. Microcomputer analysis of performance on a portable, simple visual RT task during sustained operations. *Behav Res Methods Instrum Comput.* 1985;6(17):652–655.
2. Lim J, et al. Sleep deprivation and vigilant attention. *Ann N Y Acad Sci.* 2008;1129:305–322.
3. Dorrian J, et al. Psychomotor vigilance performance: neurocognitive assay sensitive to sleep loss. In: Kushida CA, ed. *Sleep Deprivation: Clinical Issues, Pharmacology and Sleep Loss Effects.* New York, NY: Marcel Dekker; 2005: 39–70.
4. Basner M, et al. Repeated administration effects on psychomotor vigilance test performance. *Sleep.* 2018;41(1). doi:10.1093/sleep/zsx187
5. Basner M, et al. Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep.* 2011;34(5):581–591.
6. Patrick GTW, et al. On the effects of sleep loss. *Psychol Rev.* 1896;3(5):469–483.
7. Dinges DF. Probing the limits of functional capability: the effects of sleep loss on short-duration task. In: Broughton RJ, Ogilvie RD, eds. *Sleep, Arousal, and Performance.* Boston, MA: Birkhäuser; 1992: 177–188.
8. Doran SM, et al. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch Ital Biol.* 2001;139(3):1–15.
9. Lim J, et al. Imaging brain fatigue from sustained mental workload: an ASL perfusion study of the time-on-task effect. *Neuroimage* 2010;49(4):3426–3435.
10. Dinges DF, et al. Performing while sleepy: effects of experimentally-induced sleepiness. In: Monk TH, ed. *Sleep, Sleepiness and Performance.* Chichester, United Kingdom: John Wiley and Sons, Ltd.; 1991: 97–128.
11. Philip P, et al. Transport and industrial safety, how are they affected by sleepiness and sleep restriction? *Sleep Med Rev.* 2006;10(5):347–356.
12. Dinges DF. An overview of sleepiness and accidents. *J Sleep Res.* 1995;4(S2):4–14.
13. Van Dongen HP, et al. Sleep, circadian rhythms, and psychomotor vigilance. *Clin Sports Med.* 2005;24(2):237–249, vii.
14. Gunzelmann G, et al. Individual differences in sustained vigilant attention: insights from computational cognitive modeling. Paper presented at: 30th Annual Meeting of the Cognitive Science Society; July 23–26, 2008; Washington, DC.
15. Van Dongen HP, et al. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral

- functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*. 2003;26(2):117–126.
16. Gunzelmann G, et al. Fatigue in sustained attention: generalizing mechanisms for time awake to time on task. In: Ackerman PL, ed. *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications*. Washington, DC: American Psychological Association; 2010: 83–101.
 17. Jewett ME, et al. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. *Sleep*. 1999;22(2):171–179.
 18. Dinges DF, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep*. 1997;20(4):267–277.
 19. Belenky G, et al. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J Sleep Res*. 2003;12(1):1–12.
 20. Balkin TJ, et al. Comparative utility of instruments for monitoring sleepiness-related performance decrements in the operational environment. *J Sleep Res*. 2004;13(3):219–227.
 21. Mollicone DJ, et al. Response surface mapping of neurobehavioral performance: testing the feasibility of split sleep schedules for space operations. *Acta Astronaut*. 2008;63(7-10):833–840.
 22. Cohen DA, et al. Uncovering residual effects of chronic sleep loss on human performance. *Sci Transl Med*. 2010;2(14):14ra3.
 23. Van Dongen HP, et al. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep*. 2004;27(3):423–433.
 24. Neri DF, et al. Controlled breaks as a fatigue countermeasure on the flight deck. *Aviat Space Environ Med*. 2002;73(7):654–664.
 25. Van Dongen HP, et al. Caffeine eliminates psychomotor vigilance deficits from sleep inertia. *Sleep*. 2001;24(7):813–819.
 26. Czeisler CA, et al. Modafinil for excessive sleepiness associated with shift-work sleep disorder. *N Engl J Med*. 2005;353(5):476–486.
 27. Rupp TL, et al. Banking sleep: realization of benefits during subsequent sleep restriction and recovery. *Sleep*. 2009;32(3):311–321.
 28. Banks S, et al. Neurobehavioral dynamics following chronic sleep restriction: dose-response effects of one night for recovery. *Sleep*. 2010;33(8):1013–1026.
 29. Kribbs NB, et al. Effects of one night without nasal CPAP treatment on sleep and sleepiness in patients with obstructive sleep apnea. *Am Rev Respir Dis*. 1993;147(5):1162–1168.
 30. Dawson D, et al. Look before you (s)leep: evaluating the use of fatigue detection technologies within a fatigue risk management system for the road transport industry. *Sleep Med Rev*. 2014;18(2):141–152.
 31. Chua EC, et al. Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. *Sleep*. 2012;35(3):325–334.
 32. Reifman J, et al. PC-PVT 2.0: an updated platform for psychomotor vigilance task testing, analysis, prediction, and visualization. *J Neurosci Methods*. 2018;304:39–45.
 33. Basner M, et al. Development and validation of the cognition test battery. *Aerosp Med Hum Perform*. 2015;86(11):942–952.
 34. Arsintescu L, et al. Evaluation of a psychomotor vigilance task for touch screen devices. *Hum Factors*. 2017;59(4):661–670.
 35. Grandner MA, et al. Addressing the need for validation of a touchscreen psychomotor vigilance task: important considerations for sleep health research. *Sleep Health*. 2018;4(5):387–389.
 36. Basner M, et al. A new likelihood ratio metric for the psychomotor vigilance test and its sensitivity to sleep loss. *J Sleep Res*. 2015;24(6):702–713.
 37. Basner M, et al. An adaptive-duration version of the PVT accurately tracks changes in psychomotor vigilance induced by sleep restriction. *Sleep*. 2012;35(2):193–202.
 38. Basner M, et al. Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronaut*. 2011;69(11-12):949–959.