



The 3-Minute Psychomotor Vigilance Test Demonstrates Inadequate Convergent Validity Relative to the 10-Minute Psychomotor Vigilance Test Across Sleep Loss and Recovery

Caroline A. Antler, Erika M. Yamazaki, Courtney E. Casale, Tess E. Brieve and Namni Goel*

OPEN ACCESS

Edited by:

Jeanne Frances Duffy,
Brigham and Women's Hospital
and Harvard Medical School,
United States

Reviewed by:

Melissa St. Hilaire,
Brigham and Women's Hospital
and Harvard Medical School,
United States

Andrew J. K. Phillips,
Monash University, Australia

***Correspondence:**

Namni Goel
namni_goel@rush.edu

Specialty section:

This article was submitted to
Sleep and Circadian Rhythms,
a section of the journal
Frontiers in Neuroscience

Received: 15 November 2021

Accepted: 14 January 2022

Published: 15 February 2022

Citation:

Antler CA, Yamazaki EM,
Casale CE, Brieve TE and Goel N
(2022) The 3-Minute Psychomotor
Vigilance Test Demonstrates
Inadequate Convergent Validity
Relative to the 10-Minute
Psychomotor Vigilance Test Across
Sleep Loss and Recovery.
Front. Neurosci. 16:815697.
doi: 10.3389/fnins.2022.815697

Biological Rhythms Research Laboratory, Department of Psychiatry and Behavioral Sciences, Rush University Medical Center, Chicago, IL, United States

The Psychomotor Vigilance Test (PVT) is a widely used behavioral attention measure, with the 10-min (PVT-10) and 3-min (PVT-3) as two commonly used versions. The PVT-3 may be comparable to the PVT-10, though its convergent validity relative to the PVT-10 has not been explicitly assessed. For the first time, we utilized repeated measures correlation (rmcorr) to evaluate intra-individual associations between PVT-10 and PVT-3 versions across total sleep deprivation (TSD), chronic sleep restriction (SR) and multiple consecutive days of recovery. Eighty-three healthy adults (mean \pm SD, 34.7 \pm 8.9 years; 36 females) received two baseline nights (B1-B2), five SR nights (SR1-SR5), 36 h TSD, and four recovery nights (R1-R4) between sleep loss conditions. The PVT-10 and PVT-3 were completed every 2 h during wakefulness. Rmcorr compared responses on two frequently used, sensitive PVT metrics: reaction time (RT) via response speed (1/RT) and lapses (RT > 500 ms on the PVT-10 and > 355 ms on the PVT-3) by day (e.g., B2), by study phase (e.g., SR1-SR5), and by time point (1000–2000 h). PVT 1/RT correlations were generally stronger than those for lapses. The majority of correlations (48/50 [96%] for PVT lapses and 38/50 [76%] for PVT 1/RT) were values below 0.70, indicating validity issues. Overall, the PVT-3 demonstrated inadequate convergent validity with the “gold standard” PVT-10 across two different types of sleep loss and across extended recovery. Thus, the PVT-3 is not interchangeable with the PVT-10 for assessing behavioral attention performance during sleep loss based on the design of our study and the metrics we evaluated. Our results have substantial implications for design and measure selection in laboratory and applied settings, including those involving sleep deprivation.

Keywords: Psychomotor Vigilance Test, sleep deprivation, recovery, behavioral attention, repeated measures correlation, convergent validity, lapses, response speed

INTRODUCTION

One of the most commonly utilized measures in sleep research is the Psychomotor Vigilance Test (PVT), a measure of vigilant attention that requires participants to rapidly respond to visual cues randomly presented within specified interstimulus intervals (ISIs) without incorrectly responding when no stimulus is present (Dinges and Powell, 1985; Basner and Dinges, 2011). The PVT is often considered a “gold standard” measure of sleep loss deficits and it is one measure by which biomarkers or predictors of such deficits are compared (Dawson et al., 2014; Basner et al., 2015; Grandner et al., 2018; Moreno-Villanueva et al., 2018). The 10-min PVT (PVT-10) is the standard version, but more recently shorter 5-min (PVT-5) and 3-min (PVT-3) versions have been developed, particularly for applied settings that have limited time for testing (Loh et al., 2004; Basner et al., 2011).

Two published studies have directly compared performance on the PVT-10 and PVT-3 in response to sleep loss without using any other experimental manipulations: (1) the PVT-3 development study (Basner et al., 2011), which compared the PVT-10 (computer-based) and PVT-3 (handheld device-based) across total sleep deprivation (TSD) and five nights of 4 h time-in-bed (TIB) sleep restriction (SR) and (2) a validation study of smartphone-based and tablet-based 3-min PVT versions, which were compared to a laptop-based PVT-10 following 38 h TSD (Grant et al., 2017). Grant et al. (2017) reported significantly faster reaction times (RTs) and fewer lapses (PVT-10: >500 ms RT; PVT-3: >355 ms RT) on the PVT-3 relative to the PVT-10. Basner et al. (2011) also reported significantly faster RTs on the PVT-3, though they found fewer lapses on the PVT-3 only when 500 ms RT, and not 355 ms RT, was used as the lapse threshold for both PVT versions. In a study without sleep loss, Jones et al. (2018) compared performance on the PVT-10, PVT-3, and PVT-5 on the same device across 7 days in elite female basketball players and found that participants had significantly faster RTs and fewer lapses on the PVT-3 relative to the PVT-10 (and PVT-5). Additionally, a recent study involving sleep deprivation, alcohol consumption, and rest in a pressure chamber to simulate in-flight conditions compared performance on a personal computer-based PVT-10 and a handheld computer-based PVT-3 (Benderoth et al., 2021). Benderoth et al. (2021) determined that the two PVT versions had good parallel form reliability for 1/RT and lower, but still significant, correlations were found for number of lapses. Three of these studies concluded that the PVT-3 was a valid alternative to the PVT-10 (Basner et al., 2011; Grant et al., 2017; Benderoth et al., 2021), while the fourth concluded the tests were not interchangeable (Jones et al., 2018). Thus, further research is needed to systematically compare the PVT-10 and PVT-3 using the same device in highly controlled sleep loss studies.

Though averaging data from multiple time points may be necessary to meet various statistical assumptions, doing so can result in the loss of important data relating to changes in performance across time. Of note, the aforementioned sleep loss studies comparing the PVT-10 and PVT-3 (Basner et al., 2011; Grant et al., 2017) utilized averaged data in many of their

analyses, with both studies using different numbers of averaged time points, and neither study examining time-of-day variation during baseline or recovery. As a result, any information relating to discrepancies between the measures at various time points, due to possible time-of-day variation or increased homeostatic sleep pressure (Gundel et al., 2007; Fimm et al., 2015), is missing from these studies. Thus, it is important to examine individual time points to determine time-of-day variation in performance, when comparing the PVT-10 and PVT-3.

Little is known about PVT performance across extended recovery periods (e.g., more than one consecutive recovery night) following sleep deprivation (Yamazaki et al., 2021b). Some (Lamond et al., 2008; Moreno-Villanueva et al., 2018; Yamazaki et al., 2021b, 2022b) but not all studies (Wehrens et al., 2012) have demonstrated that PVT-10 performance returns to baseline levels following one night of recovery sleep after TSD. PVT-10 performance recovery following SR is more complex, with studies reporting mixed findings; these include a failure to completely return to baseline, a delayed return to baseline requiring more than one recovery night, or a return to baseline after one recovery night (Dinges et al., 1997; Banks et al., 2010; Pejovic et al., 2013; Yamazaki et al., 2021b). PVT-3 performance also returns to baseline after one night of TSD (Yamazaki et al., 2021a), but data on PVT-3 performance across recovery periods after SR are lacking. Furthermore, no studies to date have directly compared the profile of PVT-10 and PVT-3 performance across an extended recovery period of long duration (e.g., multiple consecutive nights of 12 h) after sleep loss.

Given that prior studies found significant differences between the PVT-3 and PVT-10, that no sleep loss studies administered both versions on the same device or included an extended recovery period, that most analyses utilized averaged data, and that the PVT-3 is increasingly utilized (Basner and Rubinstein, 2011; Basner et al., 2011, 2018; Hilditch et al., 2016; Grant et al., 2017; Behrens et al., 2019; Hansen et al., 2019; Yamazaki et al., 2021a), there is a significant need for studies that compare the PVT-3 to the PVT-10 on the same device in the context of different types of commonly experienced sleep loss (TSD and SR) and with an extended recovery period. Further, no study to date has evaluated the convergent validity of the PVT-3 relative to the PVT-10 while considering repeated measurements (1) across an entire sleep deprivation study, (2) across an extended recovery period, or (3) with the measures administered on the same device.

The current study utilized the repeated measures correlation (rmcorr) technique (Bland and Altman, 1995) to examine for the first time the intra-individual (within-subject) association between the PVT-10 and PVT-3 across time. This statistical method reveals the common intra-individual linear relationship, which is considered representative of the convergent validity of the measures between PVT-10 and PVT-3 metrics. Assuming the PVT-3 and PVT-10 measures are comparable in their ability to assess performance and detect change across time, it was hypothesized that relatively strong rmcorr effect sizes for comparisons between the measures for PVT lapses and PVT 1/RT would be detected, and that these relationships would remain strong regardless of time of day, since the measures should comparably capture any variations in performance due

to time effects. It was also hypothesized that all correlations would be stronger for PVT 1/RT relative to PVT lapses as well as stronger during sleep deprivation relative to baseline or recovery periods for both PVT lapses and PVT 1/RT. Lastly, it was hypothesized that correlation patterns for both PVT lapses and PVT 1/RT across the extended recovery period would not differ between those exposed to TSD versus those exposed to SR prior to recovery.

MATERIALS AND METHODS

Participants

Eighty-three healthy adults were recruited in response to study advertisements. Participants reported habitual nightly sleep durations between 6.5 and 8.5 h, with habitual bedtimes between 2200 and 0000 h, and habitual awakenings between 0600 and 0930 h; these were confirmed *via* wrist actigraphy prior to study entry. Participants did not engage in habitual napping and did not present with a sleep disorder. They did not have any acute or chronic psychological and medical conditions. Participants did not take regular medications (except for oral contraceptive use in females) and were non-smokers with body mass index values between 17.3 and 30.9 kg/m². See Yamazaki et al. (2021b) for additional details on recruitment methods, inclusion and exclusion criteria, sample characteristics, general study procedures, and participant monitoring. The protocol was approved by the University of Pennsylvania's Institutional Review Board. All participants received compensation for their participation and provided written informed consent in accordance with the Declaration of Helsinki.

Procedures

Participants engaged in a 13-day laboratory study during which they received daily checks of vital signs and symptoms by nurses (with a physician on call). The 13-day study consisted of two baseline nights (B1-B2, 10 h [2200–0800 h] and 12 h [2200–1000 h] TIB, respectively) followed by randomization to either five nights of 4 h TIB SR (SR1-SR5, 0400–0800 h, $N = 41$; Condition A) or 36 h TSD (wakefulness from 1000 to 2200 h the following day, $N = 42$; Condition B), both of which were followed by four nights of 12 h TIB (2200–1000 h) recovery sleep (R1-R4). After R1-R4, participants in the initial SR condition (Condition A) were exposed to 36 h TSD and those in the initial TSD condition (Condition B) were exposed to five nights of 4 h TIB SR. Participants were randomized in groups of four and blinded to their condition assignment until the evening after the second baseline night.

A computer-based neurobehavioral test battery was administered every 2 h during wakefulness throughout the study. Between test bouts participants were ambulatory and permitted to perform sedentary activities; however, they were not allowed to exercise. Ambient temperature was maintained between 22 and 24°C. Laboratory light levels remained constant at <50 lux during scheduled wakefulness and <1 lux during scheduled sleep periods (Yamazaki and Goel, 2020; Brieva et al., 2021; Yamazaki et al., 2021b, 2022a; Casale et al., 2022).

Neurobehavioral Measures

The computer-based neurobehavioral test battery included two widely used versions of a measure of behavioral attention: the 10-min PVT (Lim and Dinges, 2008; Basner and Dinges, 2011) and the 3-min PVT (Basner et al., 2011). Both PVT tests were administered in an environment with minimal distractions. The PVT-10 was administered before the PVT-3 during all test bouts for all participants. Participants were instructed to hit the space bar as quickly as possible after they were presented with a visual cue on the screen. Visual cues were randomly presented within specified interstimulus intervals (ISIs, or the period between the previous response and the next stimulus) specific to each measure version; the PVT-10 ISI was 2–10 s while the PVT-3 ISI was 1–4 s (Basner et al., 2011). Outcome measures were the number of lapses [RT > 500 ms on the PVT-10 and > 355 ms on the PVT-3 (Basner et al., 2011)] and response speed (mean 1/RT, henceforth referred to as 1/RT).

Statistical Analysis

Although repeated measures data are inherently valuable, their analyses can be challenging due to frequent violation of the assumptions of various statistical procedures (Keselman et al., 2001; Park et al., 2009; Bakdash and Marusich, 2017). The methods for correcting these violations, such as averaging, can result in the loss of otherwise meaningful data (Bland and Altman, 1995), and conducting analyses despite violations can result in misleading or uninterpretable results (Glass et al., 1972; Hubbard, 1978; Kenny and Judd, 1986; Scariano and Davenport, 1987). As such, instead of using Pearson's correlations, we used repeated measures correlations [rmcorr (Bakdash and Marusich, 2017; Bakdash and Marusich, 2020)], to compare PVT-10 lapses to PVT-3 lapses and to compare PVT-10 1/RT to PVT-3 1/RT. Of note, we specifically used correlational analyses because convergent validity is exclusively assessed *via* correlation (Chin and Yao, 2014). Rmcorr analyses were conducted by day (e.g., B2, SR1, R3, etc.), by study phase (e.g., SR1-SR5, R1-R4, etc.), and by time point (e.g., 1000 h, 1200 h, etc.) across the entire 13-day study and across recovery only (R1-R4) for Condition A and Condition B using the rmcorr R package (Bakdash and Marusich, 2020). By day analyses included data from the 1000–2000 h time points for B2 and for R1-R4. To retain as much data as possible, by day analyses for SR1-SR4 included early morning and late-night time points (e.g., 0800–0200 h the day after each night of SR). For SR5, only the 0800 h through 2000 h time points were collected given the start of R1 occurred immediately after SR5. TSD day was defined as 2200 h on the night of TSD through 2000 h the next day. By study phase analyses included all time points across each period (e.g., R1-R4). For Condition A and Condition B, the B2-R4 study phase included all time points from B2 through R4. The all-study days study phase included all time points from B2 through the end of TSD (2000 h) for Condition A and through the end of SR5 (2000 h) for Condition B.

Rmcorr confidence intervals (CIs) were determined using bootstrapping with replacement and using 1,000 samples (Shan et al., 2021). To meet rmcorr's linearity assumption, PVT lapses were natural log transformed [$\ln(\text{lapses} + 0.5)$]

for the by time point analyses to account for non-linear associations apparent with visual plot inspection (Cohen et al., 2003; Bakdash and Marusich, 2017). The False Discovery Rate correction of Benjamini-Hochberg (Benjamini and Hochberg, 1995) was applied to all *rmcorr* *p*-values to account for multiplicity (Gbyl et al., 2021), but notably, this did not alter the significance of any test. Thus, unadjusted *p*-values are reported. *Rmcorr* coefficient (r_{rm}) magnitude was conservatively interpreted using the following ranges: 0.00–0.29, negligible; 0.30–0.49, weak; 0.50–0.69, moderate; 0.70–0.89, strong; and 0.90–1.00, very strong (Carlson and Herdman, 2010; Mukaka, 2012; Post, 2016; Fernández-Marcos et al., 2018; Schober et al., 2018; Yadav, 2018). Furthermore, as per recommendations for interpreting convergent validity coefficients (Carlson and Herdman, 2010; Post, 2016), r_{rm} values < 0.50 indicated the PVT-3 showed inadequate convergent validity with the PVT-10, r_{rm} values > 0.70 indicated adequate convergent validity between the measures and r_{rm} values between 0.50 and 0.70 indicated validity issues between the measures. All statistical analyses were conducted in the R software environment (R Core Team, 2020). All analyses were two-sided with a *p*-value < 0.05 considered statistically significant. No participants were excluded from the analyses. Pairwise deletion was used for all analyses to minimize data loss since single data points were missing at random throughout the study; the degrees of freedom (*df*) in **Tables 1–4** serve as a proxy for the amount of data lost based on the formula $df = N(k-1) - 1$, where *N* is the total number of participants and *k* is the number of repeated measures per participant.

RESULTS

Participant Characteristics

Eighty-three healthy adults (mean ± SD, 34.7 ± 8.9 years; 36 females) (aged 21–50 years, 72.3% African American; 43.4% female) participated in the study, with *N* = 41 participants randomly assigned to Condition A (SR first) and *N* = 42 participants randomly assigned to Condition B (TSD first). There were no significant differences between conditions in age, BMI, chronotype, or the percentage of participants who were female or African American (Yamazaki et al., 2021b). There were also no significant differences between conditions in pre-study actigraphic sleep duration, onset, offset, or midpoint, or in baseline polysomnographic total sleep time or sleep onset latency (Yamazaki et al., 2021b).

Psychomotor Vigilance Test Lapses

Tables 1, 2 show r_{rm} , degrees of freedom, *p*-values, bootstrapped 95% CIs, and median and interquartile range (IQR) values for the PVT-3 and PVT-10 separately for the PVT lapses analyses. Median values were calculated for each value represented in the tables (i.e., 1000 h, B2, SR1-SR5, etc.) for all participants within each condition. We present medians, rather than means, since they are less susceptible to skewing by outliers and better reflect the central tendency of these data. Visualization is important for interpreting *rmcorr* results (Bakdash and Marusich, 2017;

Schober et al., 2018), and as such, we have included select plots (**Figures 1–3**) as examples of the range of observed effects for each analysis type.

By Study Phase

Overall, Condition B yielded stronger correlations relative to Condition A, and all the by study phase analyses for PVT lapses were significant (**Table 1**). The r_{rm} for B2-R4 was strong for Condition B and moderate for Condition A. SR1-SR5 and R1-R4 were weak for Condition A and moderate for Condition B. Interestingly, the entire study (all-study) r_{rm} was in the moderate range for both conditions. **Figure 1** presents *rmcorr* plots for the SR1-SR5 and R1-R4 analyses for Condition A and Condition B.

By Day

The by day *rmcorr* analyses revealed a wide range of *rmcorr* coefficient values for PVT lapses across study days (**Table 1**). The only correlation that was strong was R4 for Condition B. For Condition A, only R1 demonstrated a moderate correlation. For Condition B, TSD and SR1-SR3 demonstrated moderate correlations. For Condition A, weak correlations were observed for SR1-SR5 and for TSD. For Condition B, only SR4 and SR5 demonstrated weak correlations. R2 correlations were in the negligible range for both conditions while R4 was negligible for Condition A and R3 was negligible for Condition B. Neither condition demonstrated significant correlations at B2 while R3 was non-significant for Condition A and R1 was non-significant for Condition B. **Figure 2** presents B2, SR5, TSD, and R4 *rmcorr* plots for Condition A and Condition B. Notably, most individual lines approximate the overall regression line except for B2 for both conditions.

By Time Point

The entire study (all-study) duration time point *rmcorr* analyses for PVT lapses were all significant for Condition A and Condition B (**Table 2**). All r_{rm} values were moderate for Condition B, while only the 1000, 1200, and 1600 h time point correlations were moderate for Condition A (the 1800 and 2000 h time points were weak). The recovery (R1-R4) time point r_{rm} coefficients were weaker than the all-study time point coefficients. Across recovery for Condition A, the 1200 and 2000 h time point correlations were in the weak range while the 1000, 1600, and 1800 h time point correlations were in the negligible range or were non-significant. For Condition B, the all-study by time point correlations were negligible across recovery while the 1200, 1600, and 1800 h time point correlations were all non-significant. **Figure 3** presents *rmcorr* plots for 1800 h by time point analyses as an example of moderate, weak and negligible r_{rm} correlations by time point across the entire study and across recovery for both conditions.

Psychomotor Vigilance Test 1/RT

Tables 3, 4 show r_{rm} , degrees of freedom, *p*-values, bootstrapped 95% CIs, and median and IQR values for the PVT-3 and PVT-10 separately for the PVT 1/RT analyses. Median values were calculated for each value represented in the tables (i.e., 1000 h,

TABLE 1 | PVT-10 and PVT-3 lapses rmcrr results by day and by study phase.

| Condition | Statistic | Study day | | | | | | | | | | | Study phase | | | | |
|-----------|-----------------------|-----------|--------|--------|--------|---------|---------|--------|--------|--------|--------|---------|-------------|--------|---------|--------|---------|
| | | B2 | SR1 | SR2 | SR3 | SR4 | SR5 | R1 | R2 | R3 | R4 | TSD | ALL | B2-R4 | SR1-SR5 | R1-R4 | TSD |
| A | <i>r_{rm}</i> | 0.134 | 0.395 | 0.430 | 0.388 | 0.420 | 0.398 | 0.523 | 0.178 | 0.145 | 0.293 | 0.489 | 0.566 | 0.535 | 0.472 | 0.334 | 0.489 |
| | <i>df</i> | 156 | 327 | 368 | 366 | 368 | 182 | 140 | 202 | 162 | 163 | 409 | 3345 | 2894 | 1779 | 793 | 409 |
| | <i>p</i> | 0.093 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.011 | 0.065 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | CI low | -0.005 | 0.299 | 0.315 | 0.284 | 0.304 | 0.227 | 0.105 | -0.039 | -0.057 | 0.112 | 0.422 | 0.534 | 0.499 | 0.423 | 0.228 | 0.422 |
| | CI high | 0.259 | 0.483 | 0.547 | 0.489 | 0.528 | 0.516 | 0.699 | 0.350 | 0.333 | 0.443 | 0.563 | 0.597 | 0.570 | 0.520 | 0.426 | 0.563 |
| PVT-10 | Mdn | 0.00 | 1.00 | 2.00 | 2.00 | 3.00 | 3.00 | 1.00 | 1.00 | 1.00 | 1.00 | 4.00 | 2.00 | 1.00 | 2.00 | 1.00 | 4.00 |
| | (IQR) | (1.00) | (3.00) | (5.00) | (6.00) | (7.00) | (8.00) | (3.00) | (3.00) | (3.00) | (3.00) | (11.00) | (5.00) | (4.00) | (6.00) | (3.00) | (11.00) |
| PVT-3 | Mdn | 1.00 | 2.00 | 4.00 | 4.00 | 6.00 | 6.00 | 2.00 | 2.00 | 2.00 | 2.00 | 8.00 | 3.00 | 3.00 | 4.00 | 2.00 | 8.00 |
| | (IQR) | (3.00) | (7.00) | (9.00) | (9.00) | (10.00) | (14.25) | (5.00) | (6.00) | (6.00) | (6.00) | (14.00) | (9.00) | (9.00) | (10.00) | (6.00) | (14.00) |

| Condition | Statistic | B2 | TSD | R1 | R2 | R3 | R4 | SR1 | SR2 | SR3 | SR4 | SR5 | ALL | B2-R4 | TSD | R1-R4 | SR1-SR5 |
|-----------|-----------|--------|-----------------------|--------|--------|--------|--------|--------|--------|---------|---------|---------|--------|--------|---------|--------|---------|
| | | B | <i>r_{rm}</i> | 0.000 | 0.547 | 0.066 | 0.258 | 0.238 | 0.813 | 0.563 | 0.544 | 0.519 | 0.446 | 0.330 | 0.662 | 0.705 | 0.547 |
| | <i>df</i> | 149 | 388 | 166 | 209 | 167 | 167 | 335 | 377 | 377 | 377 | 186 | 3442 | 1449 | 388 | 838 | 1824 |
| | <i>p</i> | 1.000 | <0.001 | 0.392 | <0.001 | 0.002 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | CI low | -0.253 | 0.449 | -0.182 | 0.002 | -0.010 | -0.279 | 0.439 | 0.414 | 0.412 | 0.338 | 0.147 | 0.633 | 0.661 | 0.449 | 0.067 | 0.469 |
| | CI high | 0.204 | 0.637 | 0.270 | 0.473 | 0.469 | 0.909 | 0.676 | 0.648 | 0.630 | 0.547 | 0.521 | 0.692 | 0.749 | 0.637 | 0.730 | 0.577 |
| PVT-10 | Mdn | 0.00 | 2.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 2.00 | 2.00 | 2.00 | 1.00 | 0.00 | 2.00 | 0.00 | 2.00 |
| | (IQR) | (1.00) | (7.00) | (1.00) | (1.00) | (1.00) | (1.00) | (3.00) | (4.25) | (6.00) | (9.00) | (6.00) | (4.00) | (2.00) | (7.00) | (1.00) | (5.50) |
| PVT-3 | Mdn | 1.00 | 5.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 4.00 | 4.00 | 6.00 | 5.00 | 3.00 | 2.00 | 5.00 | 1.00 | 1.00 |
| | (IQR) | (3.00) | (12.00) | (3.00) | (3.00) | (3.00) | (3.00) | (8.00) | (8.00) | (10.00) | (12.00) | (11.00) | (8.00) | (5.00) | (12.00) | (3.00) | (10.00) |

PVT-10 = 10-min Psychomotor Vigilance Test; PVT-3 = 3-min Psychomotor Vigilance Test; rmcrr = repeated measures correlation; B2 = baseline day 2; SR1-SR5 = sleep restriction days 1 through 5; R1-R4 = recovery days 1 through 4; TSD = total sleep deprivation; ALL = all study days; CI = bootstrapped 95% confidence interval; Mdn = median; IQR = interquartile range.

TABLE 2 | PVT-10 and PVT-3 transformed lapses rmcrr results by time point.

| Condition | Statistic | All study days | | | | | Recovery only | | | | |
|-----------|-----------------------|----------------|--------|--------|--------|--------|---------------|--------|--------|--------|--------|
| | | 1000 h | 1200 h | 1600 h | 1800 h | 2000 h | 1000 h | 1200 h | 1600 h | 1800 h | 2000 h |
| A | <i>r_{rm}</i> | 0.594 | 0.593 | 0.523 | 0.436 | 0.461 | 0.238 | 0.319 | 0.224 | 0.148 | 0.319 |
| | <i>df</i> | 387 | 409 | 405 | 408 | 383 | 100 | 122 | 119 | 121 | 122 |
| | <i>p</i> | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.016 | <0.001 | 0.014 | 0.103 | <0.001 |
| | CI low | 0.529 | 0.525 | 0.444 | 0.360 | 0.369 | 0.057 | 0.173 | 0.047 | -0.020 | 0.126 |
| | CI high | 0.652 | 0.656 | 0.592 | 0.512 | 0.549 | 0.415 | 0.464 | 0.427 | 0.308 | 0.488 |
| PVT-10* | Mdn | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (IQR) | (4.00) | (4.00) | (4.00) | (4.00) | (4.00) | (3.00) | (3.00) | (3.00) | (3.50) | (4.00) |
| PVT-3* | Mdn | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 | 2.00 | 2.00 | 3.00 | 2.00 |
| | (IQR) | (9.00) | (8.5) | (8.00) | (8.00) | (8.00) | (5.75) | (5.00) | (5.00) | (5.00) | (6.00) |
| B | <i>r_{rm}</i> | 0.589 | 0.514 | 0.502 | 0.506 | 0.527 | 0.199 | 0.031 | 0.125 | 0.082 | 0.242 |
| | <i>df</i> | 419 | 418 | 419 | 419 | 389 | 125 | 124 | 125 | 125 | 125 |
| | <i>p</i> | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.025 | 0.733 | 0.163 | 0.359 | 0.006 |
| | CI low | 0.523 | 0.438 | 0.433 | 0.430 | 0.445 | 0.042 | -0.175 | -0.036 | -0.136 | 0.094 |
| | CI high | 0.645 | 0.577 | 0.569 | 0.574 | 0.595 | 0.346 | 0.225 | 0.299 | 0.263 | 0.391 |
| PVT-10* | Mdn | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | (IQR) | (3.00) | (2.00) | (3.00) | (3.00) | (2.00) | (1.00) | (1.00) | (1.00) | (1.00) | (1.00) |
| PVT-3* | Mdn | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | (IQR) | (7.00) | (6.00) | (6.00) | (5.75) | (6.00) | (3.00) | (3.00) | (3.00) | (3.00) | (3.00) |

PVT-10 = 10-min Psychomotor Vigilance Test; PVT-3 = 3-min Psychomotor Vigilance Test; rmcrr = repeated measures correlation; CI = bootstrapped 95% confidence interval; Mdn = median; IQR = interquartile range. * = PVT lapses median and IQR values are not transformed.

B2, SR1-SR5, etc.) for all participants within each condition. As noted in section “Psychomotor Vigilance Test Lapses,” we present medians, rather than means, since they are less susceptible to

skewing by outliers and better reflect the central tendency of these data. Select rmcrr plots as examples of the range of observed effects for each analysis type are included (Figures 4–6).

TABLE 3 | PVT-10 and PVT-3 1/RT r_{mcrr} results by day and by study phase.

| Condition | Statistic | Study day | | | | | | | | | | | Study phase | | | | |
|-----------|-----------------------|-----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------------|--------|---------|--------|--------|
| | | B2 | SR1 | SR2 | SR3 | SR4 | SR5 | R1 | R2 | R3 | R4 | TSD | ALL | B2-R4 | SR1-SR5 | R1-R4 | TSD |
| A | <i>r_{mm}</i> | 0.284 | 0.699 | 0.629 | 0.623 | 0.590 | 0.482 | 0.540 | 0.310 | 0.375 | 0.352 | 0.691 | 0.712 | 0.678 | 0.663 | 0.422 | 0.691 |
| | <i>df</i> | 156 | 327 | 368 | 366 | 368 | 182 | 140 | 202 | 162 | 163 | 409 | 3345 | 2894 | 1779 | 793 | 409 |
| | <i>p</i> | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | CI low | 0.129 | 0.627 | 0.543 | 0.539 | 0.523 | 0.328 | 0.391 | 0.146 | 0.232 | 0.189 | 0.633 | 0.690 | 0.650 | 0.632 | 0.357 | 0.633 |
| | CI high | 0.451 | 0.762 | 0.702 | 0.695 | 0.655 | 0.598 | 0.663 | 0.463 | 0.499 | 0.484 | 0.741 | 0.733 | 0.702 | 0.696 | 0.481 | 0.741 |
| PVT-10 | Mdn | 4.03 | 3.70 | 3.60 | 3.50 | 3.36 | 3.29 | 3.65 | 3.63 | 3.64 | 3.62 | 3.21 | 3.54 | 3.58 | 3.49 | 3.63 | 3.21 |
| | (IQR) | (0.75) | (0.84) | (0.80) | (0.78) | (0.81) | (0.87) | (0.85) | (0.73) | (0.69) | (0.60) | (0.80) | (0.84) | (0.83) | (0.84) | (0.72) | (0.80) |
| PVT-3 | Mdn | 4.31 | 4.01 | 3.82 | 3.76 | 3.56 | 3.59 | 4.13 | 3.95 | 4.03 | 3.92 | 3.49 | 3.85 | 3.91 | 3.76 | 4.02 | 3.49 |
| | (IQR) | (0.78) | (1.05) | (1.14) | (1.02) | (1.04) | (1.09) | (0.97) | (1.01) | (0.93) | (0.93) | (1.13) | (1.08) | (1.05) | (1.10) | (0.98) | (1.13) |

| Condition | Statistic | B2 | TSD | R1 | R2 | R3 | R4 | SR1 | SR2 | SR3 | SR4 | SR5 | ALL | B2-R4 | TSD | R1-R4 | SR1-SR5 |
|-----------|-----------|--------|-----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| | | B | <i>r_{mm}</i> | 0.305 | 0.676 | 0.408 | 0.376 | 0.370 | 0.526 | 0.747 | 0.744 | 0.691 | 0.691 | 0.609 | 0.808 | 0.784 | 0.676 |
| | <i>df</i> | 149 | 388 | 166 | 209 | 167 | 167 | 335 | 377 | 377 | 377 | 186 | 3442 | 1449 | 388 | 838 | 1824 |
| | <i>p</i> | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | CI low | 0.152 | 0.611 | 0.272 | 0.217 | 0.228 | 0.164 | 0.688 | 0.670 | 0.620 | 0.616 | 0.485 | 0.791 | 0.757 | 0.611 | 0.353 | 0.706 |
| | CI high | 0.465 | 0.734 | 0.521 | 0.498 | 0.508 | 0.695 | 0.798 | 0.807 | 0.754 | 0.757 | 0.709 | 0.823 | 0.810 | 0.734 | 0.569 | 0.767 |
| PVT-10 | Mdn | 3.88 | 3.48 | 3.89 | 3.88 | 3.90 | 3.83 | 3.60 | 3.55 | 3.50 | 3.42 | 3.45 | 3.64 | 3.78 | 3.48 | 3.88 | 3.50 |
| | (IQR) | (0.74) | (0.84) | (0.74) | (0.78) | (0.82) | (0.72) | (0.80) | (0.76) | (0.78) | (0.84) | (0.79) | (0.80) | (0.78) | (0.84) | (0.76) | (0.81) |
| PVT-3 | Mdn | 4.12 | 3.61 | 4.21 | 4.16 | 4.23 | 4.13 | 3.82 | 3.75 | 3.70 | 3.59 | 3.60 | 3.87 | 4.04 | 3.61 | 4.17 | 3.71 |
| | (IQR) | (0.97) | (1.07) | (0.98) | (0.95) | (0.98) | (0.93) | (1.03) | (0.91) | (0.99) | (1.02) | (0.95) | (1.00) | (0.95) | (1.07) | (0.95) | (0.99) |

PVT-10 = 10-min Psychomotor Vigilance Test; PVT-3 = 3-min Psychomotor Vigilance Test; 1/RT = response speed; r_{mcrr} = repeated measures correlation; B2 = baseline day 2; SR1-SR5 = sleep restriction days 1 through 5; R1-R4 = recovery days 1 through 4; TSD = total sleep deprivation; ALL = all study days; CI = bootstrapped 95% confidence interval; Mdn = median; IQR = interquartile range.

TABLE 4 | PVT-10 and PVT-3 1/RT r_{mcrr} results by time point.

| Condition | Statistic | All study days | | | | | Recovery only | | | | |
|-----------|-----------------------|----------------|--------|--------|--------|--------|---------------|--------|--------|--------|--------|
| | | 1000 h | 1200 h | 1600 h | 1800 h | 2000 h | 1000 h | 1200 h | 1600 h | 1800 h | 2000 h |
| A | <i>r_{mm}</i> | 0.748 | 0.698 | 0.697 | 0.565 | 0.651 | 0.408 | 0.463 | 0.377 | 0.296 | 0.400 |
| | <i>df</i> | 387 | 409 | 405 | 408 | 383 | 100 | 122 | 119 | 121 | 122 |
| | <i>p</i> | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | 0.001 | <0.001 |
| | CI low | 0.701 | 0.634 | 0.622 | 0.491 | 0.566 | 0.242 | 0.314 | 0.150 | 0.130 | 0.227 |
| | CI high | 0.791 | 0.759 | 0.760 | 0.636 | 0.724 | 0.578 | 0.606 | 0.570 | 0.439 | 0.564 |
| PVT-10 | Mdn | 3.61 | 3.64 | 3.65 | 3.51 | 3.54 | 3.69 | 3.73 | 3.64 | 3.55 | 3.58 |
| | (IQR) | (0.86) | (0.86) | (0.78) | (0.78) | (0.86) | (0.75) | (0.80) | (0.71) | (0.64) | (0.72) |
| PVT-3 | Mdn | 3.89 | 3.88 | 3.93 | 3.93 | 3.91 | 4.03 | 4.01 | 4.03 | 3.94 | 4.03 |
| | (IQR) | (1.05) | (1.02) | (1.05) | (1.05) | (1.14) | (0.91) | (0.91) | (1.01) | (0.90) | (1.05) |
| B | <i>r_{mm}</i> | 0.824 | 0.784 | 0.792 | 0.736 | 0.808 | 0.506 | 0.354 | 0.449 | 0.632 | 0.449 |
| | <i>df</i> | 419 | 418 | 419 | 419 | 389 | 125 | 124 | 125 | 125 | 125 |
| | <i>p</i> | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| | CI low | 0.782 | 0.740 | 0.749 | 0.682 | 0.766 | 0.356 | 0.173 | 0.265 | 0.290 | 0.321 |
| | CI high | 0.860 | 0.822 | 0.827 | 0.788 | 0.845 | 0.652 | 0.519 | 0.589 | 0.772 | 0.586 |
| PVT-10 | Mdn | 3.71 | 3.72 | 3.71 | 3.67 | 3.69 | 3.90 | 3.89 | 3.86 | 3.82 | 3.84 |
| | (IQR) | (0.81) | (0.68) | (0.78) | (0.76) | (0.71) | (0.72) | (0.71) | (0.72) | (0.80) | (0.81) |
| PVT-3 | Mdn | 3.91 | 3.94 | 3.95 | 3.97 | 3.96 | 4.15 | 4.17 | 4.20 | 4.17 | 4.12 |
| | (IQR) | (1.03) | (0.87) | (0.98) | (0.95) | (0.95) | (0.93) | (0.91) | (0.89) | (0.99) | (0.97) |

PVT-10 = 10-min Psychomotor Vigilance Test; PVT-3 = 3-min Psychomotor Vigilance Test; 1/RT = response speed; r_{mcrr} = repeated measures correlation; CI = bootstrapped 95% confidence intervals; Mdn = median; IQR = interquartile range.

By Study Phase

For PVT 1/RT r_{mcrr} analyses across study phases, Condition B generally yielded stronger correlations relative to Condition A, but all by study phase analyses were significant for both conditions (Table 3). Only the all-study days *r_{mm}* was in the strong

range for Condition A whereas the all-study days, B2-R4, and SR1-SR5 values were all in the strong range for Condition B. For Condition A, the B2-R4 and SR1-SR5 values were in the moderate range while the R1-R4 correlation was in the weak range. The R1-R4 correlation was also in the weak range for Condition B.

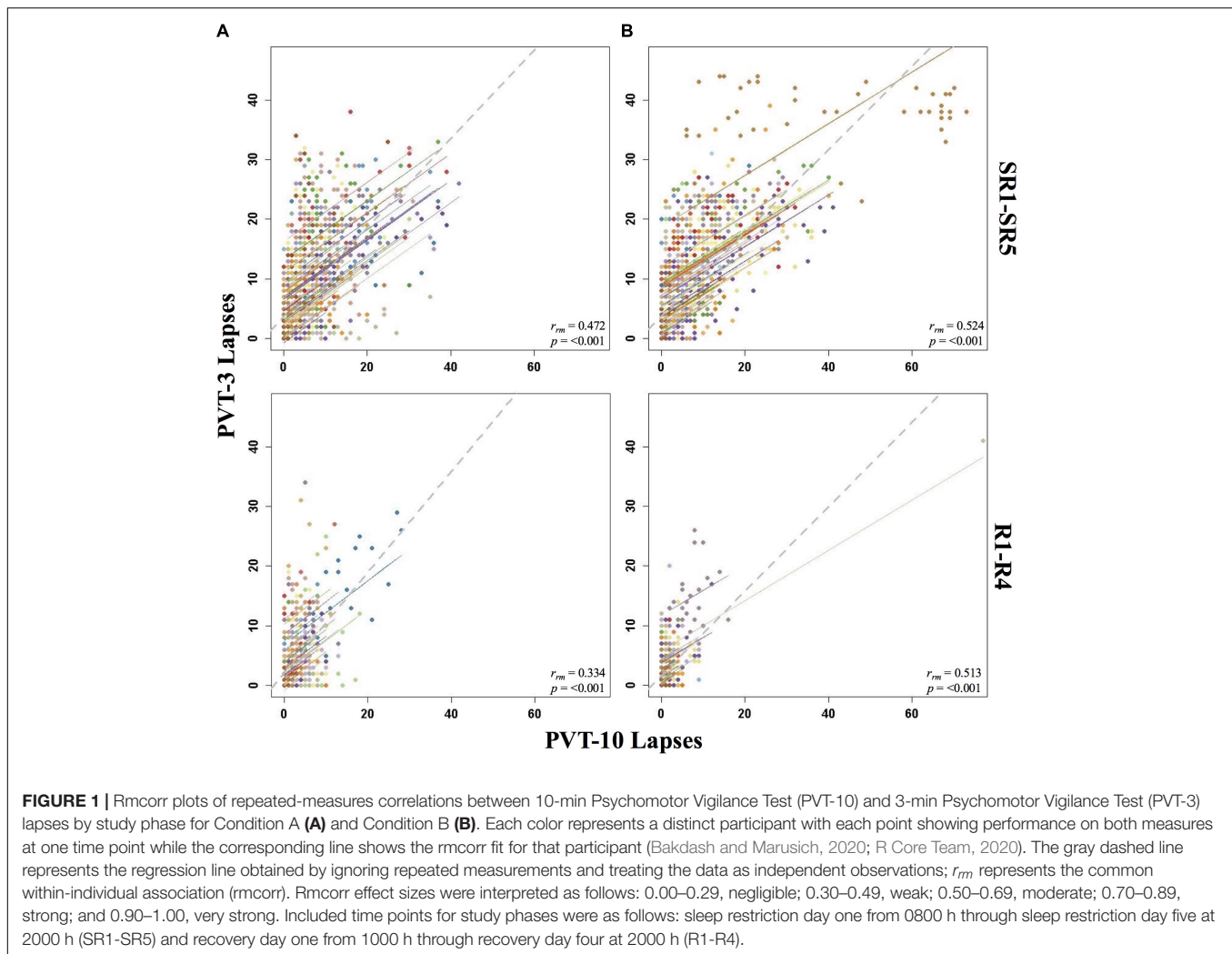


Figure 4 presents examples of rmcorr plots for the SR1-SR5 and R1-R4 analyses for Condition A and Condition B.

By Day

The by day rmcorr analyses revealed a wide range of r_{rm} values for PVT 1/RT across study days, and all correlations were significant (**Table 3**). Condition B demonstrated strong magnitude correlations for SR1 and SR2 while a strong correlation was observed for Condition A on SR1. Moderate correlations were observed for TSD in both conditions, for SR2-SR4 and R1 for Condition A, and for SR3-SR5 and R4 for Condition B. R2-R4 and R1-R3 correlations were in the weak range for Conditions A and B, respectively. The SR5 correlation also was in the weak range for Condition A and the B2 correlation was in the weak range for Condition B. The only correlation in the negligible range was B2 for Condition A. **Figure 5** presents B2, SR5, TSD, and R4 rmcorr plots for Condition A and Condition B.

By Time Point

The entire study (all-study) duration time point rmcorr analyses for PVT 1/RT were all significant for Condition A and Condition

B (**Table 4**). Every r_{rm} coefficient for the all-study analyses was strong for Condition B while the 1000, 1200, and 1600 h time points had strong correlations for Condition A [the remaining time points, 1800 and 2000 h, had moderate correlations]. The recovery (R1-R4) time point r_{rm} coefficients were weaker than the all-study time points. For Condition A, all R1-R4 time point correlations were in the weak range. For Condition B, the R1-R4 1000 and 1800 h time point correlations were in the moderate range while the 1200, 1600, and 2000 h time point correlations were in the weak range. **Figure 6** presents rmcorr plots for the 1800 h by time point analyses as an example of weak to strong magnitude rmcorr PVT 1/RT correlations by time point across the entire study and across recovery for Condition A and Condition B.

DISCUSSION

This is the first study examining the convergent validity of the PVT-3 relative to the “gold standard” PVT-10 across two commonly experienced sleep loss types followed by an

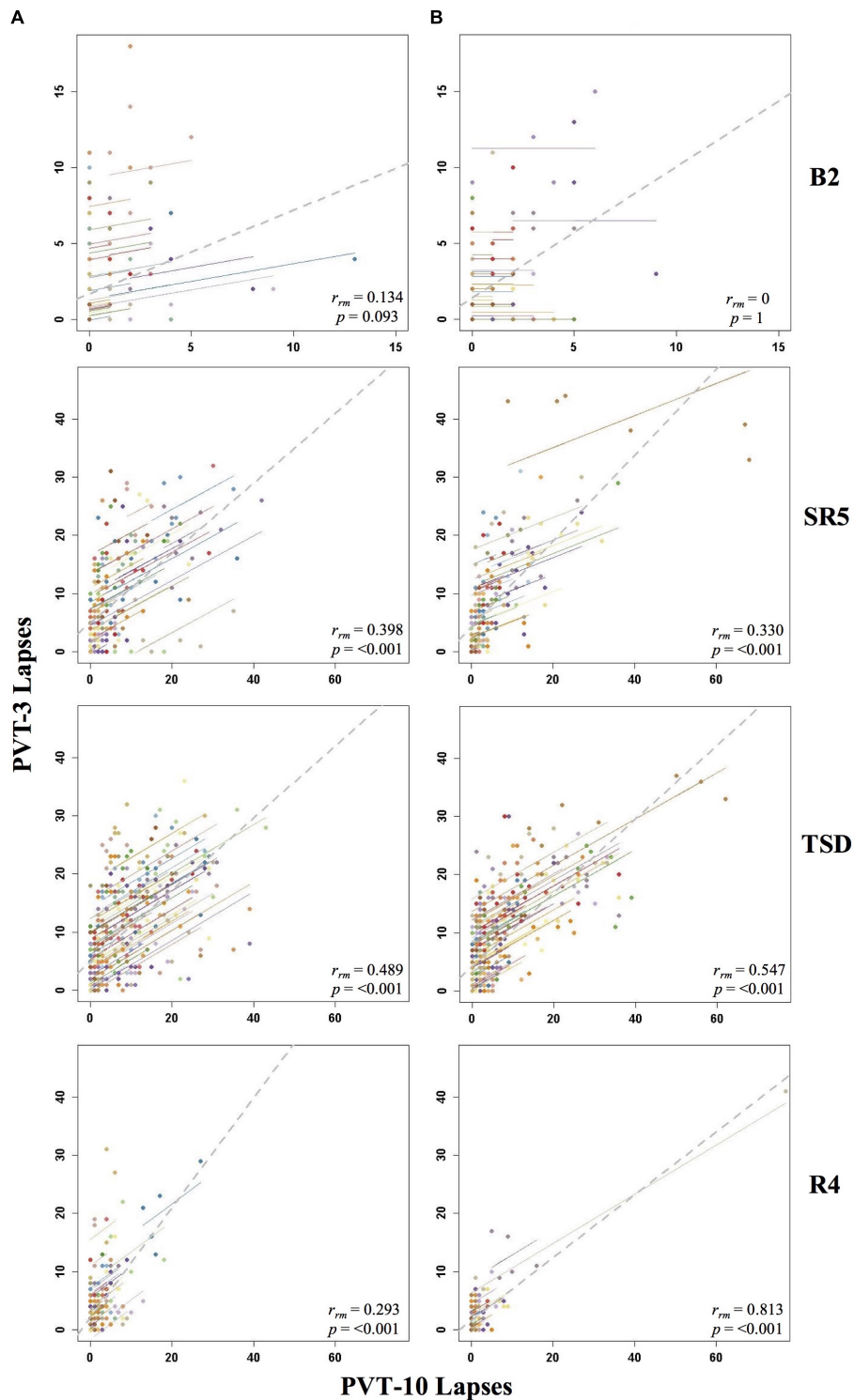
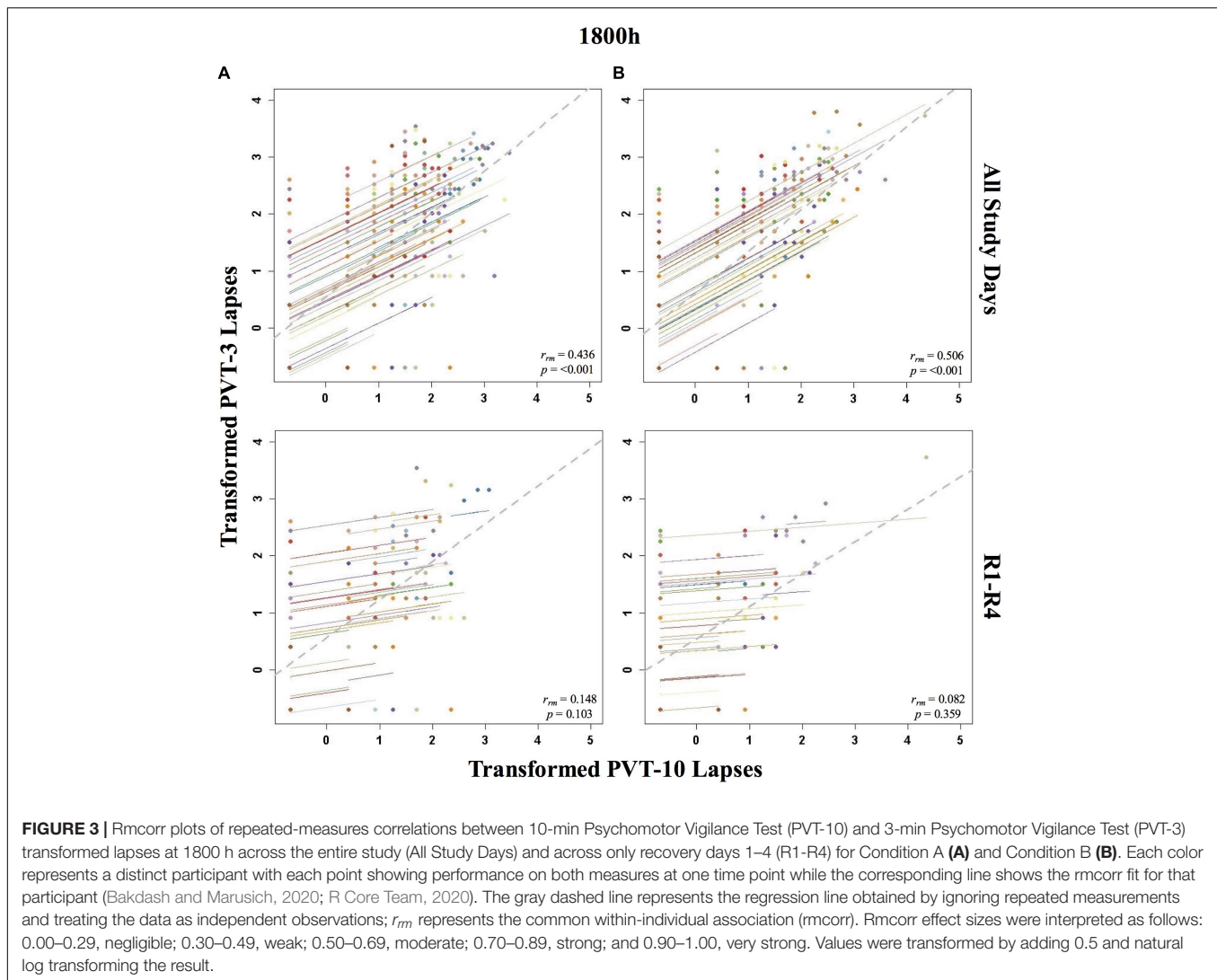


FIGURE 2 | Rmcorr plots of repeated-measures correlations between 10-min Psychomotor Vigilance Test (PVT-10) and 3-min Psychomotor Vigilance Test (PVT-3) lapses by study day for Condition A (A) and Condition B (B). Each color represents a distinct participant with each point showing performance on both measures at one time point while the corresponding line shows the rmcorr fit for that participant (Bakdash and Marusich, 2020; R Core Team, 2020). The gray dashed line represents the regression line obtained by ignoring repeated measurements and treating the data as independent observations; r_{rm} represents the common within-individual association (rmcorr). Rmcorr effect sizes were interpreted as follows: 0.00–0.29, negligible; 0.30–0.49, weak; 0.50–0.69, moderate; 0.70–0.89, strong; and 0.90–1.00, very strong. Included time points for each day were as follows: baseline day 2 (B2) from 1000 to 2200 h; sleep restriction day 5 (SR5) from 0800 to 2000 h; total sleep deprivation (TSD) from 2200 to 2000 h; and recovery day 4 (R4) from 1000 to 2000 h.

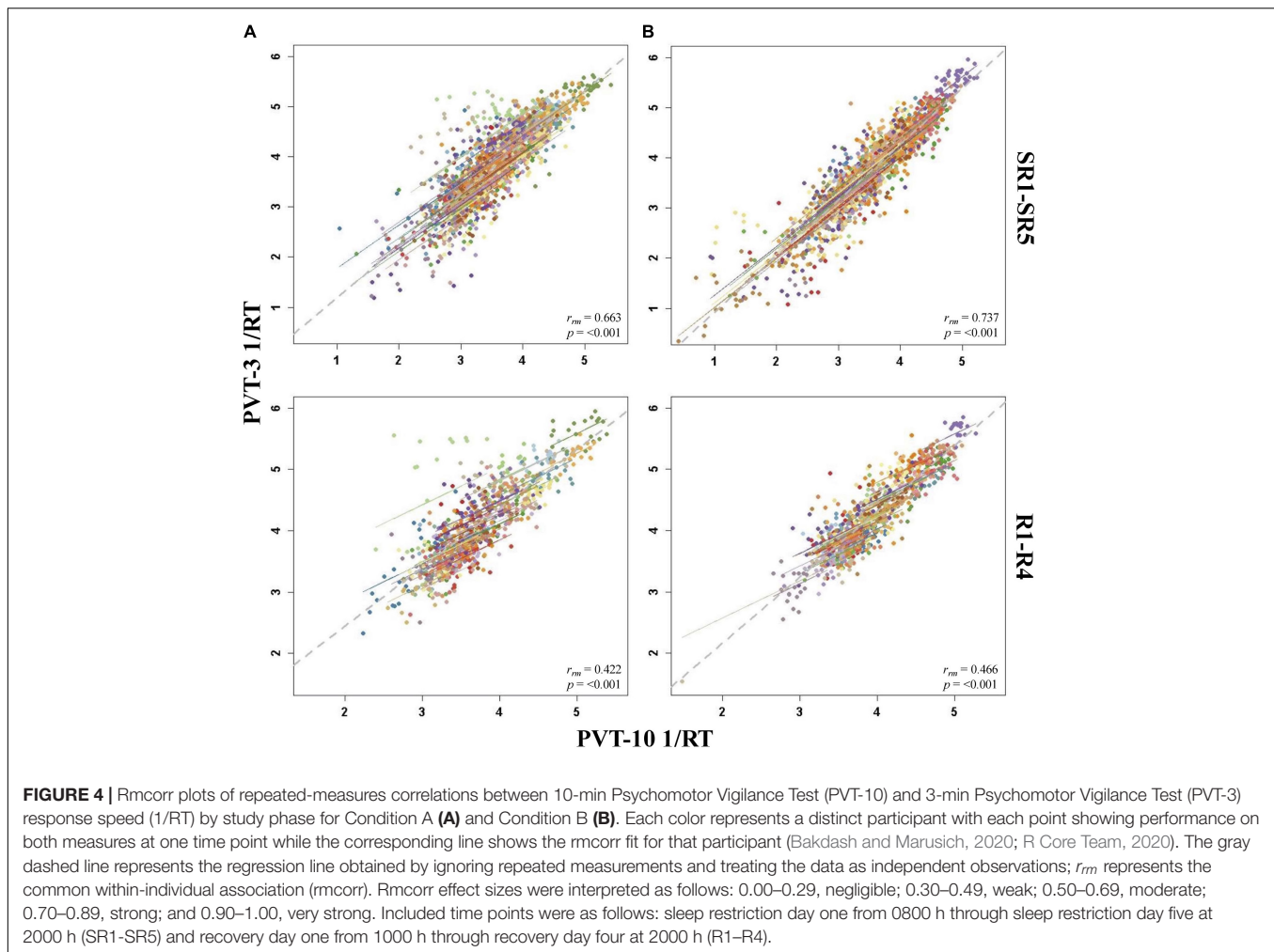


extended recovery period when administered on the same device. Correlations for PVT 1/RT were stronger relative to PVT lapses throughout the study, yet both metrics were not strongly correlated consistently throughout SR and TSD. Notably, PVT-3 lapses and 1/RT both demonstrated poor correlations with the respective PVT-10 measures during baseline and recovery periods, when participants were not undergoing experimentally induced sleep loss. Generally, the PVT-3 demonstrated inadequate convergent validity (it failed to show r_{rm} values > 0.70, strong to very strong correlations indicative of adequate convergent validity) on two frequently used PVT metrics with the “gold standard” PVT-10 across baseline, SR, TSD, and extended recovery, and when considered by individual study day, by study phases, or by specific time points.

We hypothesized that rmcrr analyses would show relatively strong correlations between the PVT-10 and PVT-3 across all study phases of the sleep deprivation study on two frequently utilized PVT outcome metrics. Contrary to our expectations, only two r_{rm} coefficient values (out of 50; 4%) were above 0.70

for PVT lapses (both of those occurred in Condition B) while only 12 r_{rm} coefficient values (24%) were above 0.70 for PVT 1/RT (ten of those occurred in Condition B). Examined more granularly by time point, only analyses including all-study days for PVT 1/RT had r_{rm} values above 0.70, with Condition A only having one time point (1000 h) above this value, while no recovery time point analyses had r_{rm} values above 0.32 for PVT lapses or r_{rm} values above 0.63 for PVT 1/RT. Given that convergent validity coefficients less than 0.70 (less than strong or very strong) indicate validity issues (Carlson and Herdman, 2010; Post, 2016), these findings suggest that the convergent validity of the PVT-3 compared with the PVT-10 is inadequate based on two commonly used outcome metrics. Notably, our results are in line with, and expand upon, the findings of Jones et al. (2018) who concluded that the measures were not interchangeable.

Considered across the study, correlations for PVT 1/RT were generally stronger than those for PVT lapses, thus supporting our hypothesis. Our results correspond with previous studies that found PVT lapses more consistently differed and demonstrated



lower correlations between the PVT-10 and PVT-3 relative to PVT 1/RT (Basner et al., 2011; Grant et al., 2017; Benderoth et al., 2021). Of note, although Grant et al. (2017) and Benderoth et al. (2021) found significant correlations between the PVT-3 and PVT-10 for 1/RT and lapses, the majority of correlations for lapses were below 0.70 and therefore similarly suggestive of validity issues. Interestingly, Basner et al. (2011) and Grant et al. (2017) attributed the observed differences between the measures to the use of different devices for administration [which notably is also similar to the Benderoth et al. (2021) study], yet the tests were administered on the same device in the present study [and in Jones et al. (2018)]; thus, the use of distinct devices does not likely explain any observed measure differences. The difference in lapse thresholds between the two PVT versions may also have potentially influenced the weaker correlations. While the thresholds we utilized (>500 ms for the PVT-10 and >355 ms for the PVT-3) are the most widely accepted thresholds in the sleep and circadian field, it could be valuable to investigate whether different lapse thresholds may better capture consistencies between the PVT-10 and the PVT-3 in future work.

Our hypothesis that correlations would be strongest during sleep deprived study phases compared to rested study phases

(baseline and recovery) was generally supported. Correlations for recovery only time points, across recovery days, and across the recovery study phase were almost always weaker than those demonstrated during periods of sleep deprivation, while baseline correlations were non-significant, negligible, or weak. Since the Jones et al. (2018) study included multiple days of repeated measurements without sleep deprivation and determined the PVT-10 and PVT-3 measures did not perform comparably, our results provide further evidence that the lack of comparability is true for baseline measurements, and our results extend these findings to recovery periods following sleep loss interventions.

Different ISI durations between PVT versions could have potentially contributed to differential sensitivity to performance during rested and sleep deprived periods. One study failed to find a differential impact on PVT-10 and PVT-3 performance but reported that TSD enhanced the ISI effect (Yang et al., 2018). Since no study has evaluated the impact of differing ISIs on PVT performance in SR, it is unclear if SR would produce similar effects. If a differential impact of SR on the ISI effect were to be demonstrated, that factor, in tandem with incomplete recovery of PVT performance following SR (Yamazaki et al., 2021b) could explain the generally stronger correlations observed for those

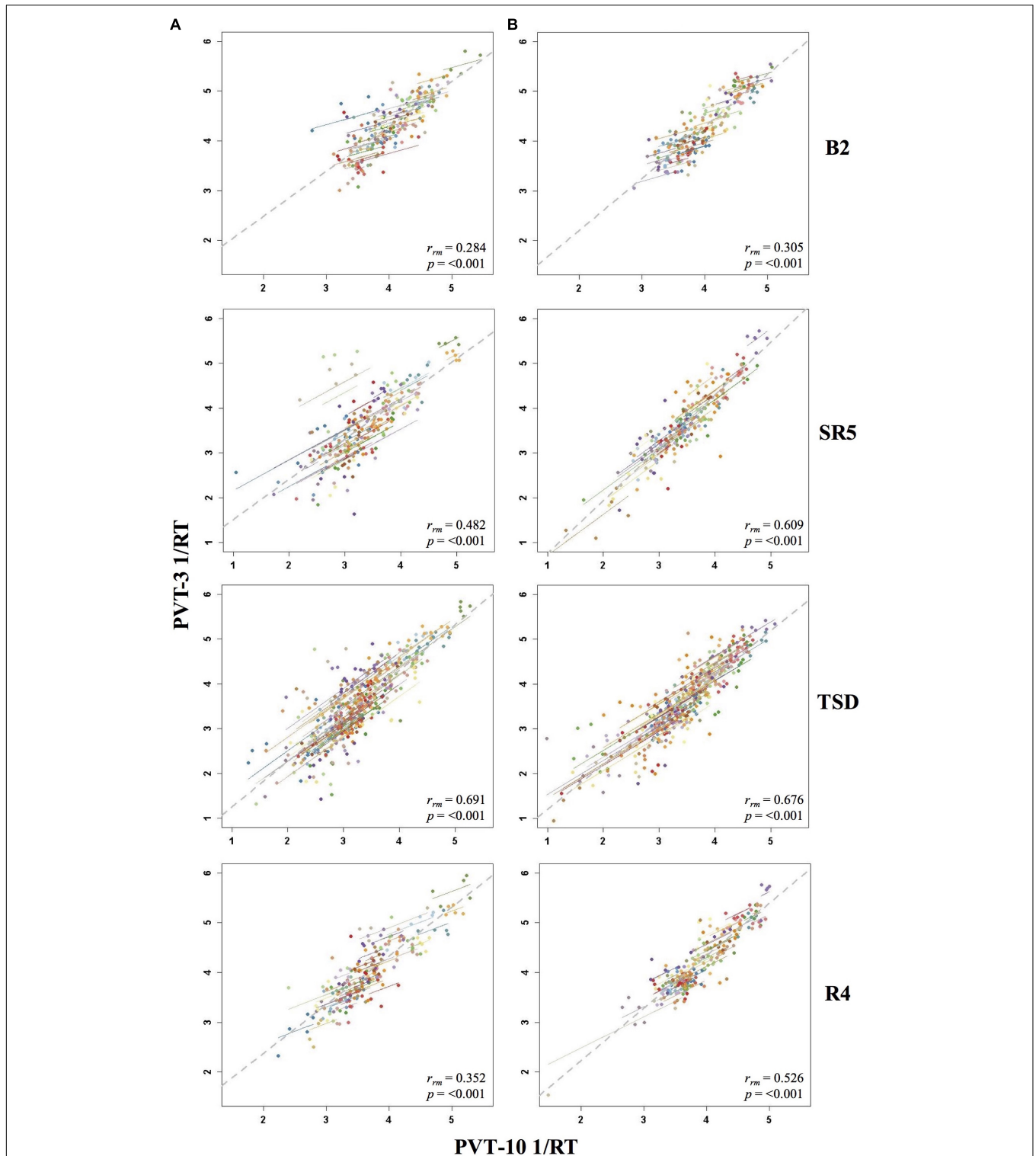
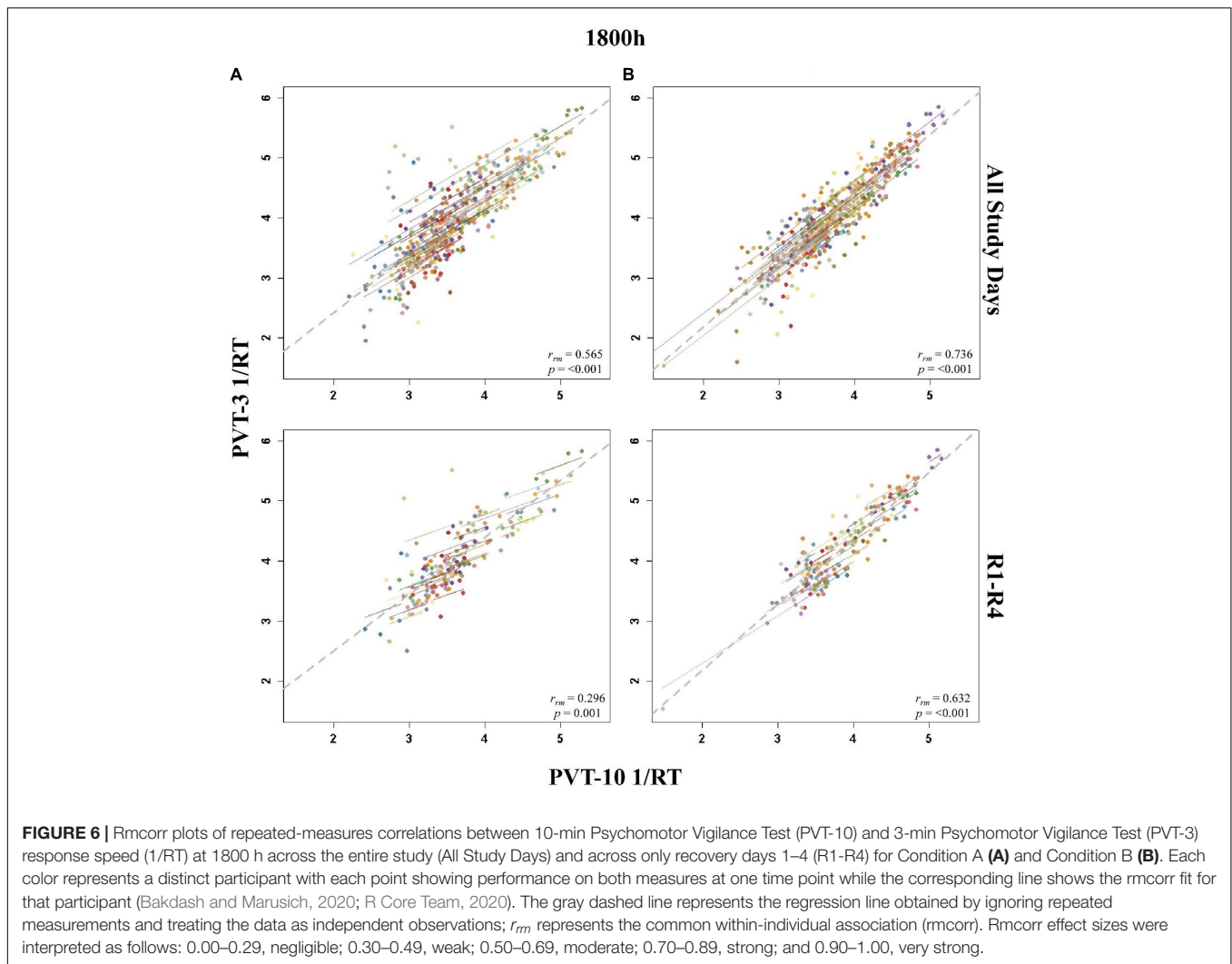


FIGURE 5 | Rmcorr plots of repeated-measures correlations between 10-min Psychomotor Vigilance Test (PVT-10) and 3-min Psychomotor Vigilance Test (PVT-3) response speed (1/RT) by study day for Condition A (A) and Condition B (B). Each color represents a distinct participant with each point showing performance on both measures at one time point while the corresponding line shows the rmcorr fit for that participant (Bakdash and Marusich, 2020; R Core Team, 2020). The gray dashed line represents the regression line obtained by ignoring repeated measurements and treating the data as independent observations; r_{rm} represents the common within-individual association (rmcorr). Rmcorr effect sizes were interpreted as follows: 0.00–0.29, negligible; 0.30–0.49, weak; 0.50–0.69, moderate; 0.70–0.89, strong; and 0.90–1.00, very strong. Included time points for each day were as follows: baseline day 2 (B2) from 1000 to 2200 h; sleep restriction day 5 (SR5) from 0800 to 2000 h; total sleep deprivation (TSD) from 2200 to 2000 h; and recovery day 4 (R4) from 1000 to 2000 h.



exposed to TSD first (Condition B) relative to those exposed to SR first (Condition A). Future research on the potential differential impact of SR relative to TSD on the ISI effect, the severity of incomplete recovery from SR, and the effects of sleep loss exposure order on neurobehavioral performance may provide insight into how these various factors interact.

The fact that the PVT-10 and PVT-3 are not comparable in their ability to measure behavioral attention during rested periods has implications for study design, test selection, and use of biomarkers or predictors relating to performance (Dawson et al., 2014; Basner et al., 2015; Grandner et al., 2018; Moreno-Villanueva et al., 2018). Moreover, studies that utilize baseline PVT-3 lapses or 1/RT to evaluate change with sleep loss should be interpreted with caution since reliance on such metrics for baseline comparisons is likely to yield misleading results. Given the importance of individual differences in neurobehavioral performance in sleep research (Leproult et al., 2003; Van Dongen et al., 2004; Tucker et al., 2007; Van Dongen, 2012; Spaeth et al., 2014; Ramakrishnan et al., 2015; Rusterholz et al., 2016; Dennis et al., 2017; Goel, 2017; Tkachenko and Dinges, 2018;

Hajali et al., 2019; Letzen et al., 2019; Yamazaki and Goel, 2020; Brieva et al., 2021; Casale et al., 2022; Yamazaki et al., 2022a), it is essential that studies evaluating change over time, including in neurobehavioral response to sleep loss, are able to accurately assess baseline performance (Glymour et al., 2005). If PVT-3 and PVT-10 lapses and/or 1/RT outcome metrics are different when individuals are not impaired yet are slightly more comparable during periods of sleep deprivation, then it is possible (if not likely) that studies evaluating change over time that include baseline or recovery measurements would find disparate results depending on these measures.

Lastly, our hypothesis that correlation patterns across the extended recovery study phase would not differ between those exposed to TSD relative to those exposed to SR was not supported. Interestingly, correlations between the measures on R1 for those exposed to SR first (Condition A) were moderate while they were non-significant and negligible to weak for those exposed to TSD first (Condition B). Correlations on R2–R3 were comparable between conditions, yet R4 demonstrated negligible to weak correlations for Condition A, but these were

moderate-to-strong for Condition B. Given that research on performance throughout extended recovery periods is limited, interpretation of these findings is challenging, but might relate to a differential recovery neurobehavioral performance profile following SR relative to TSD. Indeed, this is in line with a study from Yamazaki et al. (2021b) that found PVT-10 impairments resulting from SR did not fully recover even after extended recovery (four 12 h TIB nights), yet impairments fully recovered following TSD (Yamazaki et al., 2021a). Further studies are needed to explore these sleep loss-recovery dynamics.

Our study had a few limitations. These findings may not apply to different duration versions of the PVT such as the PVT-5 (Loh et al., 2004; Roach et al., 2006; Arsintescu et al., 2019) or to handheld versions of the PVT (Lamond et al., 2008). It is also possible that test administration order (i.e., in our study the PVT-10 always occurred before the PVT-3) may have impacted response of one or both measures, including potential time-on-task effects, though one study did not find evidence to support such an order effect (Basner et al., 2011). Future studies should examine the potential influence of order of test administration when assessing the convergent validity of PVT versions on the same device, as previous studies, including our own, did not examine this important possibility. In addition, our study did not assess system latency bias (the delay in response time due to test administration hardware and software platforms), which might have impacted our results, particularly for PVT lapses (Basner et al., 2021). Because the PVT-10 and PVT-3 were administered during the same testing session and on the same device during each test bout, any impact such bias had on PVT 1/RT should have been comparable between measures.

The highly controlled nature of our study, the large sample, the same device administration and the ability to utilize all available study data are unique strengths in the context of an evaluation of convergent validity. In our study, although PVT 1/RT specifically demonstrated relatively strong correlations across SR and TSD, most correlations were below an acceptable threshold for the measures to be considered interchangeable, while such correlations for PVT lapses were consistently below that threshold. Our results, coupled with the discordant results between the PVT-10 and PVT-3 observed during sleep loss on at least one major outcome metric in prior PVT-3 studies (Basner et al., 2011; Grant et al., 2017; Jones et al., 2018), indicate that the PVT-3 and PVT-10 may be measuring explicitly different constructs even during periods of impaired functioning. Thus, based on our experimental protocol and findings, the PVT-3 should be interpreted with caution when compared with the PVT-10 for lapses and 1/RT metrics. Since the PVT-3 has been proposed as a test to capture sleep loss-induced deficits rapidly and reliably in applied settings, such as aviation and other transportation sectors (Dawson et al., 2014), hospital shift work (Behrens et al., 2019), and security-related situations (Basner and Rubinstein, 2011), it is critical to discern whether the shorter 3-min version is a valid assessment of vigilant attention under both sleep-deprived and rested conditions. If so, this would allow for a rapid evaluation of objective vigilant attention in various operational settings. Other shorter alternatives to the PVT-10 such as the 5-min PVT version (Loh et al., 2004; Roach

et al., 2006; Arsintescu et al., 2019) are available for use in applied settings, yet Jones et al. (2018) found differences between 5- and 10-min PVT; as such, future studies including those involving sleep loss in applied settings are needed. This work could evince the utility of time-efficient objective metrics in assessing an individual's level of vigilant attention impairment from sleep loss.

DATA AVAILABILITY STATEMENT

The data generated and analyzed during the current study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Pennsylvania's Institutional Review Board. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NG designed the original study during which the analyzed data were collected and extracted and retains oversight and control of all data as Principal Investigator. NG provided financial support, including all research grant support, all personnel support, and all publication fees. CA conceived of the present study, identified and verified the statistical methods, and performed the analyses. EY, CC, and TB checked analysis scripts, plots, and results for accuracy. CA wrote the manuscript. All authors contributed to the interpretation of the results, provided critical feedback, helped shape the research, analysis, and manuscript, and reviewed and approved the final manuscript.

FUNDING

This work was primarily supported by the Department of the Navy, Office of Naval Research (Award No. N00014-11-1-0361) to NG Other support provided by National Aeronautics and Space Administration (NASA) NNX14AN49G and grant 80NSSC20K0243 (to NG), National Institutes of Health grant NIH R01DK117488 (to NG) and Clinical and Translational Research Center grant UL1TR000003. None of the sponsors had any role in the following: design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

ACKNOWLEDGMENTS

We thank the faculty and staff of the Unit for Experimental Psychiatry for their contributions to this study in terms of data collection.

REFERENCES

- Arsintescu, L., Kato, K. H., Cravalho, P. F., Feick, N. H., Stone, L. S., and Flynn-Evans, E. E. (2019). Validation of a touchscreen psychomotor vigilance task. *Accid. Anal. Prev.* 126, 173–176. doi: 10.1016/j.aap.2017.1.041
- Bakdash, J. Z., and Marusich, L. R. (2017). Repeated measures correlation. *Front. Psychol.* 8:456. doi: 10.3389/fpsyg.2017.00456
- Bakdash, J. Z., and Marusich, L. R. (2020). *rmcorr: Repeated Measures Correlation. R Package Version 0.4.1*. Available Online at: <https://CRAN.R-project.org/package=rmcorr>. (accessed April 16, 2021).
- Banks, S., Van Dongen, H. P. A., Maislin, G., and Dinges, D. F. (2010). Neurobehavioral dynamics following chronic sleep restriction: dose-response effects of one night for recovery. *Sleep* 33, 1013–1026. doi: 10.1093/sleep/33.8.1013
- Basner, M., and Dinges, D. F. (2011). Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss. *Sleep* 34, 581–591. doi: 10.1093/sleep/34.5.581
- Basner, M., Hermosillo, E., Nasrini, J., McGuire, S., Saxena, S., Moore, T. M., et al. (2018). Repeated administration effects on psychomotor vigilance test performance. *Sleep* 41:zsx187. doi: 10.1093/sleep/zsx187
- Basner, M., McGuire, S., Goel, N., Rao, H., and Dinges, D. F. (2015). A new likelihood ratio metric for the psychomotor vigilance test and its sensitivity to sleep loss. *J. Sleep Res.* 24, 702–713. doi: 10.1111/jsr.12322
- Basner, M., Mollicone, D., and Dinges, D. F. (2011). Validity and sensitivity of a brief psychomotor vigilance test (PVT-B) to total and partial sleep deprivation. *Acta Astronaut.* 69, 949–959. doi: 10.1016/j.actaastro.2011.07.015
- Basner, M., Moore, T. M., Nasrini, J., Gur, R. C., and Dinges, D. F. (2021). Response speed measurements on the psychomotor vigilance test: how precise is precise enough? *Sleep* 44:zsaa121. doi: 10.1093/sleep/zsaa121
- Basner, M., and Rubinstein, J. (2011). Fitness for duty. *J. Occup. Environ. Med.* 53, 1146–1154. doi: 10.1097/jom.0b013e31822b8356
- Behrens, T., Burek, K., Pallapies, D., Kösters, L., Lehnert, M., Beine, A., et al. (2019). Decreased psychomotor vigilance of female shift workers after working night shifts. *PLoS One* 14:e0219087. doi: 10.1371/journal.pone.0219087
- Benderoth, S., Hörmann, H.-J., Schießl, C., and Elmenhorst, E.-M. (2021). Reliability and validity of a 3-min psychomotor vigilance task in assessing sensitivity to sleep loss and alcohol: fitness for duty in aviation and transportation. *Sleep* 44:zsab151. doi: 10.1093/sleep/zsab151
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bland, J. M., and Altman, D. G. (1995). Statistics notes: calculating correlation coefficients with repeated observations: part 1—correlation within subjects. *BMJ* 310:446. doi: 10.1136/bmj.310.6977.446
- Brieva, T. E., Casale, C. E., Yamazaki, E. M., Antler, C. A., and Goel, N. (2021). Cognitive throughput and working memory raw scores consistently differentiate resilient and vulnerable groups to sleep loss. *Sleep* 44:zsab197. doi: 10.1093/sleep/zsab197
- Carlson, K. D., and Herdman, A. O. (2010). Understanding the impact of convergent validity on research results. *Org. Res. Methods* 15, 17–32. doi: 10.1177/1094428110392383
- Casale, C. E., Yamazaki, E. M., Brieva, T. E., Antler, C. A., and Goel, N. (2022). Raw scores on subjective sleepiness, fatigue, and vigor metrics consistently define resilience and vulnerability to sleep loss. *Sleep* 45:zsab228. doi: 10.1093/sleep/zsab228
- Chin, C. L., and Yao, G. (2014). “Convergent validity,” in *Encyclopedia of Quality of Life and Well-Being Research*, ed. A. C. Michalos (Dordrecht: Springer), 1275–76. doi: 10.1007/978-94-007-0753-5_573
- Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. New York: Routledge.
- Dawson, D., Searle, A. K., and Paterson, J. L. (2014). 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.* 18, 141–152. doi: 10.1016/j.smr.2013.03.003
- Dennis, L. E., Wohl, R. J., Selame, L. A., and Goel, N. (2017). Healthy adults display long-term trait-like neurobehavioral resilience and vulnerability to sleep loss. *Sci. Rep.* 7:14889. doi: 10.1038/s41598-017-14006-7
- Dinges, D. F., Pack, F., Williams, K., Gillen, K. A., Powell, J. W., Ott, G. E., et al. (1997). Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. *Sleep* 20, 267–277. doi: 10.1093/sleep/20.4.267
- Dinges, D. F., and Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav. Res. Methods Instrum. Comput.* 17, 652–655. doi: 10.3758/bf03200977
- Fernández-Marcos, T., de la Fuente, C., and Santacreu, J. (2018). Test-retest reliability and convergent validity of attention measures. *Appl. Neuropsychol. Adult* 25, 464–472. doi: 10.1080/23279095.2017.1329145
- Fimm, B., Brand, T., and Spijkers, W. (2015). Time-of-day variation of visuo-spatial attention. *Br. J. Psychol.* 107, 299–321. doi: 10.1111/bjop.12143
- Gbyl, K., Støttrup, M. M., Raghava, J. M., Jie, S. X., and Videbeck, P. (2021). Hippocampal volume and memory impairment after electroconvulsive therapy in patients with depression. *Acta Psychiatr. Scand.* 143, 238–252. doi: 10.1111/acps.13259
- Glass, G. V., Peckham, P. D., and Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Rev. Educ. Res.* 42, 237–288. doi: 10.3102/00346543042003237
- Glymour, M. M., Weuve, J., Berkman, L. F., Kawachi, I., and Robins, J. M. (2005). When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am. J. Epidemiol.* 162, 267–278. doi: 10.1093/aje/kwi187
- Goel, N. (2017). Neurobehavioral effects and biomarkers of sleep loss in healthy adults. *Curr. Neurol. Neurosci. Rep.* 17:89. doi: 10.1007/s11910-017-0799-x
- Grandner, M. A., Watson, N. F., Kay, M., Ocaño, D., and Kientz, J. A. (2018). Addressing the need for validation of a touchscreen psychomotor vigilance task: important considerations for sleep health research. *Sleep Health* 4, 387–389. doi: 10.1016/j.sleh.2018.08.003
- Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., and 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. *Behav. Res. Methods* 49, 1020–1029. doi: 10.3758/s13428-016-0763-8
- Gundel, A., Marsalek, K., and Radu, C. (2007). Sleep-related and time-of-day variations in fatigue and psychomotor performance. *Somnologie (Berl)* 11, 186–191. doi: 10.1007/s11818-007-0305-9
- Hajali, V., Andersen, M. L., Negah, S. S., and Sheibani, V. (2019). Sex differences in sleep and sleep loss-induced cognitive deficits: the influence of gonadal hormones. *Horm. Behav.* 108, 50–61. doi: 10.1016/j.yhbeh.2018.12.013
- Hansen, D. A., Layton, M. E., Riedy, S. M., and Van Dongen, H. P. (2019). Psychomotor vigilance impairment during total sleep deprivation is exacerbated in sleep-onset insomnia. *Nat. Sci. Sleep* 11, 401–410. doi: 10.2147/nss.s224641
- Hilditch, C. J., Centofanti, S. A., Dorrian, J., and Banks, S. (2016). A 30-minute, but not a 10-minute nighttime nap is associated with sleep inertia. *Sleep* 39, 675–685. doi: 10.5665/sleep.5550
- Hubbard, R. (1978). The probable consequences of violating the normality assumption in parametric statistical analysis. *Area* 10, 393–398.
- Jones, M. J., Dunican, I. C., Murray, K., Peeling, P., Dawson, B., Halson, S., et al. (2018). The psychomotor vigilance test: a comparison of different test durations in elite athletes. *J. Sports Sci.* 36, 2033–2037. doi: 10.1080/02640414.2018.1433443
- Kenny, D. A., and Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychol. Bull.* 99, 422–431. doi: 10.1037/0033-2909.99.3.422
- Keselman, H. J., Algina, J., and Kowalchuk, R. K. (2001). The analysis of repeated measures designs: a review. *Br. J. Math. Stat. Psychol.* 54, 1–20. doi: 10.1348/000711001159357
- Lamond, N., Jay, S. M., Dorrian, J., Ferguson, S. A., Jones, C., and Dawson, D. (2008). The dynamics of neurobehavioural recovery following sleep loss. *J. Sleep Res.* 16, 33–41. doi: 10.1111/j.1365-2869.2007.00574.x
- Leproult, R., Colecchia, E. F., Berardi, A. M., Stickgold, R., Kosslyn, S. M., and Van Cauter, E. (2003). Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 284, R280–R290. doi: 10.1152/ajpregu.00197.2002
- Letzen, J. E., Remeniuk, B., Smith, M. T., Irwin, M. R., Finan, P. H., and Seminowicz, D. A. (2019). Individual differences in pain sensitivity are associated with cognitive network functional connectivity following one night

- of experimental sleep disruption. *Hum. Brain Mapp.* 41, 581–593. doi: 10.1002/hbm.24824
- Lim, J., and Dinges, D. F. (2008). Sleep deprivation and vigilant attention. *Ann. N. Y. Acad. Sci.* 1129, 305–322. doi: 10.1196/annals.1417.002
- Loh, S., Lamond, N., Dorrian, J., Roach, G., and Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behav. Res. Methods Instrum. Comput.* 36, 339–346. doi: 10.3758/bf03195580
- Moreno-Villanueva, M., Scheven, G. V., Feiveson, A., Bürkle, A., Wu, H., and Goel, N. (2018). The degree of radiation-induced DNA strand breaks is altered by acute sleep deprivation and psychological stress and is associated with cognitive performance in humans. *Sleep* 41:zsy067. doi: 10.1093/sleep/zsy067
- Mukaka, M. M. (2012). Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med. J.* 24, 69–71.
- Park, E., Cho, M., and Ki, C.-S. (2009). Correct use of repeated measures analysis of variance. *Ann. Lab. Med.* 29, 1–9. doi: 10.3343/kjlm.2009.29.1.1
- Pejovic, S., Basta, M., Vgontzas, A. N., Kritikou, I., Shaffer, M. L., Tsaoussoglou, M., et al. (2013). Effects of recovery sleep after one work week of mild sleep restriction on interleukin-6 and cortisol secretion and daytime sleepiness and performance. *Am. J. Physiol. Endocrinol. Metab.* 305, E890–E896. doi: 10.1152/ajpendo.00301.2013
- Post, M. W. (2016). What to do with “moderate” reliability and validity coefficients? *Arch. Phys. Med. Rehabil.* 97, 1051–1052. doi: 10.1016/j.apmr.2016.04.001
- R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ramakrishnan, S., Lu, W., Laxminarayan, S., Wesensten, N. J., Rupp, T. L., Balkin, T. J., et al. (2015). Can a mathematical model predict an individual’s trait-like response to both total and partial sleep loss? *J. Sleep Res.* 24, 262–269. doi: 10.1111/jsr.12272
- Roach, G. D., Dawson, D., and Lamond, N. (2006). Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiol. Int.* 23, 1379–1387. doi: 10.1080/07420520601067931
- Rusterholz, T., Tarokh, L., Van Dongen, H. P. A., and Achermann, P. (2016). Interindividual differences in the dynamics of the homeostatic process are trait-like and distinct for sleep versus wakefulness. *J. Sleep Res.* 26, 171–178. doi: 10.1111/jsr.12483
- Scariano, S. M., and Davenport, J. M. (1987). The effects of violations of independence assumptions in the one-way ANOVA. *Am. Stat.* 41, 123–129. doi: 10.2307/2684223
- Schober, P., Boer, C., and Schwarte, L. A. (2018). Correlation coefficients. *Anesth. Analg.* 126, 1763–1768. doi: 10.1213/ane.0000000000002864
- Shan, G., Zhang, H., and Barbour, J. (2021). Bootstrap confidence intervals for correlation between continuous repeated measures. *Stat. Methods Appl.* 30, 1175–1195. doi: 10.1007/s10260-020-00555-1
- Spaeth, A. M., Dinges, D. F., and Goel, N. (2014). Sex and race differences in caloric intake during sleep restriction in healthy adults. *Am. J. Clin. Nutr.* 100, 559–566. doi: 10.3945/ajcn.114.086579
- Tkachenko, O., and Dinges, D. F. (2018). Interindividual variability in neurobehavioral response to sleep loss: a comprehensive review. *Neurosci. Biobehav. Rev.* 89, 29–48. doi: 10.1016/j.neubiorev.2018.03.017
- Tucker, A. M., Dinges, D. F., and Van Dongen, H. P. A. (2007). Trait interindividual differences in the sleep physiology of healthy young adults. *J. Sleep Res.* 16, 170–180. doi: 10.1111/j.1365-2869.2007.00594.x
- Van Dongen, H. P. (2012). Connecting the dots: from trait vulnerability during total sleep deprivation to individual differences in cumulative impairment during sustained sleep restriction. *Sleep* 35, 1031–1033.
- Van Dongen, H. P., Baynard, M. D., Maislin, G., and Dinges, D. F. (2004). Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep* 27, 423–433. doi: 10.1093/sleep/27.3.423
- Wehrens, S. M. T., Hampton, S. M., Kerkhofs, M., and Skene, D. J. (2012). Mood, alertness, and performance in response to sleep deprivation and recovery sleep in experienced shiftworkers versus non-shiftworkers. *Chronobiol. Int.* 29, 537–548. doi: 10.3109/07420528.2012.675258
- Yadav, S. (2018). Correlation analysis in biological studies. *J. Pract. Cardiovasc. Sci.* 4:116. doi: 10.4103/jpcs.jpcs_31_18
- Yamazaki, E. M., Antler, C. A., Casale, C. E., MacMullen, L. E., Ecker, A. J., and Goel, N. (2021a). Cortisol and C-reactive protein vary during sleep loss and recovery but are not markers of neurobehavioral resilience. *Front. Physiol.* 12:782860. doi: 10.3389/fphys.2021.782860
- Yamazaki, E. M., Antler, C. A., Lasek, C. R., and Goel, N. (2021b). Residual, differential neurobehavioral deficits linger after multiple recovery nights following chronic sleep restriction or acute total sleep deprivation. *Sleep* 44:zsaa224. doi: 10.1093/sleep/zsaa224
- Yamazaki, E. M., Casale, C. E., Brieva, T. E., Antler, C. A., and Goel, N. (2022a). Concordance of multiple methods to define resiliency and vulnerability to sleep loss depends on psychomotor vigilance test metric. *Sleep* 45:zsab249. doi: 10.1093/sleep/zsab249
- Yamazaki, E. M., Rosendahl-Garcia, K. M., Casale, C. E., MacMullen, L. E., Ecker, A. J., Kirkpatrick, J. N., et al. (2022b). Left ventricular ejection time measured by echocardiography differentiates neurobehavioral resilience and vulnerability to sleep loss and stress. *Front. Physiol.* 12:795321. doi: 10.3389/fphys.2021.795321
- Yamazaki, E. M., and Goel, N. (2020). Robust stability of trait-like vulnerability or resilience to common types of sleep deprivation in a large sample of adults. *Sleep* 43:zsz292. doi: 10.1093/sleep/zsz292
- Yang, F. N., Xu, S., Chai, Y., Basner, M., Dinges, D. F., and Rao, H. (2018). Sleep deprivation enhances inter-stimulus interval effect on vigilant attention performance. *Sleep* 41:zsy189. doi: 10.1093/sleep/zsy189

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Antler, Yamazaki, Casale, Brieva and Goel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.