The impact of sleep loss on performance monitoring and error-monitoring: A systematic review and meta-analysis

Authors

Johanna M. Boardman[1]

Kate Porcheret[123]

Jacob W. Clark[1]

Thomas Andrillon[1]

Anna W.T. Cai[1]

Clare Anderson[1]

Sean P.A Drummond[1]

1. Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Australia.
2. Norwegian Centre for Violence and Traumatic Stress Studies, Norway.
3. Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway.

Corresponding Author
Sean P.A Drummond
sean.drummond@monash.edu
18 Innovation Walk, Room 542
Clayton, VIC 3800

Funding

JMB is supported by a scholarship provided by Defence Science & Technology Group.

This review was also supported by a grant from the Office of Naval Research Global: Award N629091712142

Conflicts of interest

The authors have no conflicts of interest to declare.

Summary

Awareness of performance deficits and errors during sleep loss could be protective against the consequences of sleep deprivation, however, it is unclear whether sleep deprived individuals have insight into their performance. We conducted a systematic review and meta-analysis of the impact of sleep loss (sleep duration <6 hours) on monitoring of performance and errors using Embase, MEDLINE, PsycINFO & Cochrane Central. We identified 28 studies, 11 of which were appropriate for meta-analysis. The systematic review indicated limited consensus regarding sleep loss impacts on performance monitoring, due to substantial differences in study methodology. However, participants typically demonstrated more conservative estimates of performance during sleep loss. Error-monitoring literature was more consistent, indicating an impairment in error-monitoring following sleep loss. Meta-analyses supported the findings of the systematic review. In terms of methodology, we found the performance monitoring literature is limited by an overreliance on correlational designs, which are likely confounded by response bias. The error-monitoring literature is limited by very few studies utilising behavioural measures to directly measure error-awareness. Future performance monitoring studies must employ methods which control for confounds such as bias, and error-monitoring studies must incorporate combined behavioural and ERP measures to better understand the impact of sleep loss on error-monitoring.

Keywords: metacognition, sleep deprivation, sleep restriction, self-monitoring, error-detection, error-awareness, awareness, cognition

Glossary of terms

Calibration	The extent to which the proportion of correct responses is reflected in confidence ratings. Also see *"metacognitive sensitivity"*.
Calibration curve	A plot of the proportion of correct responses associated with each confidence level.
Error-awareness	The conscious recognition an error has occurred.
Error correction	The act of changing an incorrect response to a correct response.
Error-detection	The brain's unconscious response to an error.
Error-related negativity	A negative deflection occurring 80-150ms after an erroneous response.
Error-positivity	A positive deflection occurring 200-400ms after an erroneous response.
Event-related potential	An electrical potential evoked in response to a specific event, such as a stimulus or response.
Metacognition	Awareness and understanding of one's own thought processes
Meta-*d'*	A bias free measure of how well confidence ratings track task accuracy.
Metacognitive bias	The extent to which an individual is predisposed to report low or high confidence.
Metacognitive sensitivity	The extent to which confidence ratings distinguish between correct and incorrect responses.
Post-error adjustment	When performance on the trial following an error improves.
Signal Detection Theory	Theory regarding how stimuli are detected under different conditions.

Abbreviations

EEG - electroencephalography

EPP – error-preceding positivity

ERN – error-related negativity

ERP – event-related potential

Pe – error positivity

PVT – psychomotor vigilance task

RT – Reaction time

SR – sleep restriction

TSD – total sleep deprivation

VAS – visual analogue scale

Introduction

Many individuals are required to function on limited sleep due to work demands (e.g., shift workers, first responders, and military personnel). As a result, these individuals are at risk of making errors due to sleepiness, which may have dangerous consequences. For example, drowsy driving more than triples the odds of hazardous driving events [1] and up to 20% of motor vehicle accidents are attributed to sleep-related impairment [2, 3]. Furthermore, recent estimates suggest 29% of adults report errors at work because of sleepiness [4]. Finding solutions to combat the adverse consequences of sleep loss is therefore critical. Engaging in compensatory behaviours such as napping [5], using stimulants, or adapting the way a task is performed, are commonly employed strategies to minimise sleep-related risk. Compensatory behaviours may be anticipatory, occurring before task performance when we expect decrements due to sleep loss, or current, occurring "in the moment", when we feel our performance is sub-optimal. However, to engage in these compensatory mechanisms, we must first recognise or anticipate we currently are, or are likely to be, impaired. Performance monitoring and error-monitoring are types of "metacognition", which refers to knowledge of our own thoughts and behaviours. Error-monitoring can be further considered a subtype of performance monitoring. However, to aid clarity for the purpose of this review, we use the terms "performance monitoring" and "error-monitoring" to refer to two distinct sets of studies, as described below. Within the context of sleep, effective monitoring is critical for an individual to recognise, in the moment, the negative effects of sleep loss on behaviour.

Performance monitoring

Performance monitoring is the self-assessment of performance on a particular cognitive task or activity (i.e., driving). The concept of performance monitoring has been extensively studied outside of the sleep context, particularly in the field of psychophysics and through the Signal Detection Theory framework [6]. Within performance monitoring, an individual is typically required to assess their performance immediately before or after the task [7-9], by providing a subjective estimate of performance as a percentage of correct responses or a standardised Likert scale (e.g., “poor”, “excellent”). These subjective responses are subsequently compared with objective performance outcomes. Metacognitive theory posits higher confidence in performance should coincide with higher accuracy – a concept known as metacognitive sensitivity [10]. As such, some studies assess participants’ subjective confidence they have provided a correct answer after each trial [11-13]. Confidence is then correlated with trial accuracy to establish the degree of metacognitive sensitivity. Similarly, metacognitive studies may also use a “calibration curve” to examine confidence-accuracy relationships where the proportion of correct responses associated with each confidence level is analysed [11, 14]. Good calibration is represented by an increase in the proportion of correct responses with higher confidence and represents ideal metacognitive sensitivity. In contrast, a stable proportion of correct responses across all possible confidence ratings reflects poor calibration [11]. These calibration curves can also evidence biases towards over-confidence or under-confidence.

Error-monitoring

Error-monitoring focuses specifically on the detection and recognition of incorrect responses. The error monitoring system comprises three components 1) error detection - the unconscious response of the brain to an error; 2) error awareness – the conscious recognition an error occurred, and; 3) post error adjustments - where errors are corrected, and/or where cognitive performance improves immediately following an error. Improvements in trials following errors may be reflected in post-error slowing of reaction time (RT) (reflecting more cautious processing of stimuli), or greater accuracy following error trials relative to following correct responses [15, 16]. A common method to examine error-monitoring, is the recording of event-related potentials (ERPs), which reflect stimulus or response-locked synchronous activations of large neuronal assemblies [17]. In the context of error-monitoring, ERPs in response to correct and incorrect responses are compared. The error-related negativity (ERN), a negative deflection occurring 80 to100ms following an erroneous response is an index of error detection, evident on error trials, regardless of whether an individual recognises an error has occurred [18, 19]. In contrast, the error-positivity (Pe), a positive deflection occurring 200 to 400ms after an erroneous response, is a measure of error awareness, appearing larger, or only, on trials a participant deems incorrect [20, 21]. The ERN is thought to reflect neural activity in the anterior cingulate cortex [22] while the Pe is suggested to have prefrontal/parietal origins [23]. Amplitude of ERN and Pe is posited to reflect the processing of an error, or the significance of a mismatch between an intended versus executed outcome, whereas latency reflects the timing of this process [24, 25].

Review aims and objectives

While studies on sleep and performance monitoring emerged almost thirty years ago, a comprehensive review examining multiple methods of performance monitoring and sleep has not been conducted. Such a review will provide much needed clarity on whether sleep loss impairs our ability to detect deficits in our performance, and what constitutes the best method to address this question. In contrast to performance monitoring, the sleep and error-monitoring literature is still largely in its infancy, although research outside the sleep field has unravelled the core neural mechanisms underlying error-monitoring [19, 20, 26]. A review of the literature aimed at sleep researchers is therefore timely, providing valuable insight into what we have so far uncovered and where future research should now focus. Therefore, this review aims to 1) summarise the existing sleep, performance and error monitoring literature; 2) identify key methodological similarities and differences among studies; 3) assess the methodological quality of the current literature; and 4) identify limitations and future directions. Additionally, while we intended to conduct a meta-analysis of this literature, relatively few manuscripts provided the necessary data to calculate a standardised effect size. This meta-analysis was therefore performed in support of our primary aims.

Methods

Search method and studies selection

Four databases (Embase, MEDLINE, PsycINFO & Cochrane Central) were initially searched in October 2019, with an updated search conducted in September 2020. A combination of subject headings and key terms were used. Subject headings for the main concepts of sleep deprivation and performance monitoring were identified. Key terms relating to sleep loss, such as “sleep deprivation”, and “sleep disturbance” were used in conjunction with key words relating to performance monitoring (e.g., “self-assessment”, “meta-cognition”) and error monitoring (“error awareness”, “error detection”). These terms could appear anywhere in the manuscript. No limits were applied on publication date. See Supp Fig1 for details of full search strategy. To ensure an accurate search strategy had been formulated, a Gold-Set of 21 articles were tested against the search. Grey literature was also sought from Open Access Theses & Dissertations. Further relevant studies were obtained by searching the reference list of retrieved papers. Works not written in English were not included in the review. Findings are reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards [27].

Inclusion and exclusion criteria

Titles and abstracts were imported into Covidence [28] where they were independently screened for relevance by two researchers (JMB & KP). Commonly, within the sleep literature, a given paper may report performance monitoring analyses, but not report those results in the abstract. Therefore, to maximise the likelihood relevant articles would be included, any study which included a manipulation of sleep duration, with an objective performance outcome, was included at this initial stage. At the full text stage, performance monitoring studies were included if: 1) there was a measurable partial (sleep duration <6 hours) or total sleep deprivation (TSD) manipulation; 2) an objective performance outcome, and; 3) a measure of subjective performance or confidence, and; 4) a direct statistical comparison between objective and subjective measures was conducted. Error monitoring studies were included if : 1) there was a measurable partial (sleep duration <6 hours) or TSD manipulation, and; 2) a measure of error monitoring with ERPs alone, or in combination with behavioural data, or; 3) a measure of error detection with ERPs, or; 4) a measure of post-error adjustments with behavioural data, or ERPs alone, or a combination of both, or; 4) a measure of error correction with behavioural data or ERPs alone, or a combination of both, or; 5) a measure of error awareness with ERPs or in conjunction with behavioural data.

Studies were excluded if: 1) the manuscript was a review; 2) study participants were animals; 3) study participants were children or adolescents; 4) there was no subjective performance measure; 5) objective measures were not statistically compared with subjective measures; 6) correlations or difference scores were not examined as a result of a sleep measure or manipulation. In addition, while some studies have compared measures of subjective sleepiness ratings to objective performance and interpreted this as “performance monitoring”, we conceptualise assessment of sleepiness and subjective assessment of performance as different constructs. Thus, we did not include subjective sleepiness studies in this review.

Conference abstracts which addressed the research questions were included, even if they did not report statistics. In cases where we could not determine if conference abstracts and peer-reviewed papers were from the same dataset, the abstract was excluded. When multiple abstracts appeared to include data from the same dataset, the version with most detailed results reported was selected for inclusion. Risk of bias was completed with the JBI Critical Appraisal Checklist for Quasi-Experimental Studies [29].

Data extraction

For each eligible study in the systematic review, data were extracted for authors, year of publication, mean age, sample size, performance measures, study design, and whether the authors reported a significant effect of the sleep manipulation on: 1) objective performance; 2) performance monitoring ability/error-monitoring and; 3) subjective sleepiness. Additional information extracted from error-monitoring studies included electrode site where ERPs were measured, number of electrodes for overall EEG setup, and time windows used to define ERP components of interest. For the meta-analysis of the performance monitoring literature, means and standard deviations were extracted for objective performance and subjective performance, whereas for error-monitoring, means and standard deviations were extracted for ERN and Pe amplitude.

Quantitative analysis

Those performance monitoring papers for which relevant data could be extracted were included for the meta-analysis. Separate sub-group analyses were performed to investigate effects of condition (control [i.e., well-rested] and experimental [i.e., sleep disrupted]) and task type (working memory, attention, exam, and physical performance) on subjective versus objective performance. Condition x task type interaction effects were also assessed. For ERN and Pe amplitude, separate meta-analyses were performed for those papers where data were available. For papers which did not report means and standard deviations for objective and subjective performance, or Pe or ERN amplitude, authors were contacted to obtain raw data. Authors were given one month to respond to data requests. In cases where data were provided in graphs, Engauge Digitizer [30] was used to extract raw data directly from figures.

Meta-analyses were performed using RevMan Version 5.4 (The Cochrane Collaboration, Copenhagen, Denmark). Effect sizes were pooled using a random effects model [31]. Standardised mean difference between studies was assessed in the analysis of objective versus subjective performance. Interaction effects for these data were analysed using GraphPad Prism 8.3.0 (California, USA). Interaction effects were followed-up with post-hoc Sidak's multiple comparisons tests. For ERN or Pe amplitude, weighted mean difference between studies was investigated. Heterogeneity between effect sizes was quantified using the I^2 statistic and evaluated with an alpha level of $p < 0.05$ [31]. An I^2 value of 0% indicates no observed heterogeneity, whereas larger values indicate increasing heterogeneity (i.e., 25%, 50%, 75% = low, medium, and high levels of heterogeneity, respectively). The presence of publication bias was evaluated via visual inspection of funnel plots [32].

Results

Results of the database search are provided in Fig1 and a summary of included article characteristics is provided in Supp. Table 1. After full screening, 28 studies were included: 16 addressing performance monitoring, nine addressing error-monitoring, and three addressing both. Eighteen studies had within-subjects designs, nine were between-subjects designs, and one was a mixed design. Twenty employed TSD protocols and eight employed sleep restriction (SR) protocols. A large range of cognitive tasks were assessed in the performance monitoring literature, with most measuring memory or sustained attention. In contrast, almost all error-monitoring studies measured response inhibition. Twelve studies controlled for sleep duration prior to conducting the experiment, ten of which monitored sleep/wake patterns (8 x diary, 6 x actigraphy). The impact of sleep loss on performance monitoring and error-monitoring outcomes is summarised in Tables 1-2.

Performance monitoring

Three questions were addressed by the performance monitoring literature; 1) Does sleep loss influence performance monitoring accuracy? (eleven studies: [7-9, 12, 14, 33-38]) 2) What is the relationship between objective performance and subjective ratings of performance during sleep loss? (four studies: [39-42]) and 3) What is the impact of sleep loss on the relationship between objective accuracy and confidence? (Six studies: [11-14, 43, 44]).

Performance monitoring accuracy

Of the eleven studies addressing the question of whether sleep loss influences performance monitoring accuracy, ten of these studies employed TSD protocols while two employed SR protocols. Across studies, the findings are mixed.

Seven studies reported no effect of sleep loss on performance monitoring accuracy [7-9, 14, 33, 34, 36]. Of these, one found an overestimation of performance [36], two reported an underestimation of performance [33, 34], and one reported accurate performance monitoring [9]. In a series of studies, Baranski et al. [7, 8, 14], found performance monitoring accuracy differed between tasks, with general knowledge performance overestimated, mental addition performance underestimated in one study and accurate in another, and perceptual comparison performance accurately monitored in two studies, although only up to 54 hours TSD. However Boardman et al. [9] found accurate performance monitoring in a moderately difficult serial subtraction task. In all studies, accurate performance monitoring and over- or underestimations of performance did not significantly change with sleep loss. In contrast, seven studies report an effect of sleep loss on performance monitoring accuracy [7-9, 12, 35, 37, 38]. Specifically, Boardman et al. [9]and Blagrove & Akehurst [12] observed a greater underestimation of performance (i.e., impaired performance monitoring) during TSD on a difficult working memory, and logical reasoning task, but an increase in performance monitoring accuracy following TSD on the Psychomotor Vigilance Task (PVT) and an abstract reasoning task. Baranski et al. [7] also found task differences such that participants underestimated logical reasoning performance and overestimated vigilance performance to a greater extent following sleep loss. Additionally Baranski & Pigaeu [8]found overestimation of perceptual comparison performance after 54 hours of TSD. Terlizzese et al. [37] found sleep restricted females underestimated their performance relative to females in a control group, but the same effect was not observed in males. Similarly, Daviaux et al. [38] reported a greater underestimation of stepping height capability in sleep deprived individuals compared to a well-rested control group. In contrast, Lust [35], who obtained trial by trial estimates of performance, found SR did not affect correct trial identification, but sleep restricted individuals were more likely to incorrectly identify an error as correct. Thus, SR was associated with impaired performance monitoring, whereby individuals overestimated performance.

Overall, across the eleven accuracy studies, seven reported a significant change in performance monitoring accuracy after sleep loss. Of 14 tasks across seven studies, six tasks showed greater underestimation of performance, five tasks showed greater overestimation of performance, and three tasks showed more accurate performance monitoring when sleep deprived.

Quantitative analysis – performance monitoring accuracy

Seven studies were included in the meta-analysis [9, 12, 33, 34, 37, 38, 45], with one to two studies for each cognitive domain. Results of the overall model showed underestimation of performance ($I^2 = 82\%$, $p < .00001$). The condition x cognitive domain interaction was marginally significant ($F(3, 1623) = 3.40$, $p = 0.05$) such that only tests of attention differed significantly according to condition (e.g., experimental groups underestimated performance less than control groups; $p < 0.05$). The sub-group analysis of condition was marginally significant, with all participants underestimating performance, and underestimation was smaller, when sleep deprived ($I^2 = 69.70$, $p = 0.07$; Supp. Fig2). Sub-group analysis of task type demonstrated greater underestimation of performance during tests of attention, followed by exam performance, and then working memory ($I^2 = 77.50$, $p < .01$; Supp. Fig3). Of note, one study [34] included in the meta-analysis reported separate objective data for congruent and incongruent trial types, but an overall subjective performance score. An overall objective performance score could not be calculated with the data provided, and therefore the same subjective performance score was used for both congruent and incongruent trial types. While this study still provides valuable information relevant to the research question, we acknowledge these data are not as accurately addressing the research question as the other data included. To assess the impact of including Hsieh et al.,[34] on the overall analysis, we reran the analysis excluding the study and found no substantial changes in effect size or significance. A visual inspection of funnel plots (Supp. Figs4-5) found no evidence of clear publication bias.

Subjective ratings of performance

Four studies examined the relationship between objective performance and subjective ratings of performance [39-42]. Subjective ratings were typically measured with a Visual Analogue Scale (VAS) and all studies utilised SR protocols (one night of 2-4 hours of sleep). Kosmadopoulos et al. [40] reported subjective ratings of performance reflected actual decreases in objective performance during sleep loss, although correlations between objective and subjective performance were reduced during sleep loss. Philip et al. [41] found a significant correlation between subjective ratings and objective performance in a laboratory well-rested condition, but not in well-rested and SR on-road driving conditions. Biggs et al. [39] reported no significant correlations between measures either while well-rested or during SR and Smith et al. [42] had similar results, finding; 1) Discrepancies between objective and subjective performance were largest in sleep restricted individuals, and; 2) participants who performed the most poorly showed the most accurate self-assessments [42]. Interestingly, in all studies, regardless of the significance or direction of correlations, subjective ratings of performance changed in the expected direction following sleep loss.

Confidence-accuracy studies

Six studies examined the relationship between accuracy and confidence, five under TSD conditions [11-14, 43], one under chronic SR [44]. Two studies examined confidence with calibration curve analyses [11, 14], reporting in both studies calibration was not influenced by TSD. However, participants typically demonstrated overconfidence on general knowledge but were well-calibrated on perceptual comparison and mental addition. Three studies examined confidence-accuracy correlations. Both Aidman et al. [43] and Blagrove et al. [12] report strong correlations between accuracy and confidence during well-rested test sessions, while TSD led to a decrease in confidence-accuracy . Similarly, Harrison & Horne [13] and Matthew et al. [44] reported TSD and SR led to more false alarms where participants were more likely to report higher confidence for incorrect responses.

Overall, calibration analyses show no influence of TSD, while correlational studies suggest TSD may weaken the relationship between confidence and accuracy. It is also possible correlations became weaker because of ceiling or floor effects on confidence ratings during TSD. Regardless, the overall set of findings suggest under TSD conditions, high confidence does not necessarily equate to more accurate performance.

Error-monitoring

ERN and Pe Amplitude

Ten studies [33, 35, 45-52] examined amplitude changes in the error-related negativity (unconscious error detection) and nine [33, 35, 45-50, 52] investigated the error positivity (conscious error awareness). Six of ten studies reported reductions in ERN amplitude following sleep loss [33, 47, 48, 50-52]. Interestingly, one study found no impact of sleep loss on amplitude for participants who received monetary incentives for good performance, but amplitude was reduced for those who did not receive an incentive [33]. Five of nine studies reported reductions in Pe amplitude following sleep loss [33, 45, 47, 49, 52]. Conversely, Renn & Cote [50] reported an increase in Pe amplitude following sleep loss when trials were collapsed across error block. An additional study examined changes in the error-preceding positivity (EPP), suggested to index a reduction in attention prior to an error [53]. The EPP was observed only in the well-rested condition, suggesting limited error-related attention following TSD [53].

Quantitative analysis – ERN & Pe Amplitude

Seven studies were included in the meta-analysis [33, 45, 48-52]. Results demonstrated reduced amplitude of both ERN (amplitude less negative in sleep deprived groups (I^2 = 71%, $p < 0.0001$; Fig2) and Pe (amplitude less positive in sleep deprived groups) (I^2 = 8%, $p < 0.001$; Fig3). A visual inspection of funnel plots (Supp. Fig6-7) revealed no clear evidence of publication bias among these studies. As two studies [33, 48] utilised methodologies which may have influenced the effect of sleep loss on error-monitoring, we reran this analysis with these studies removed. Results demonstrated reduced amplitude of both ERN (Z= 5.2) and Pe (Z=3.01).

ERN and Pe Latency

Five studies examined ERN and/or Pe latency under sleep loss conditions [33, 47, 49, 50, 52]. Of the four ERN studies, one reported increased ERN latency [33] with sleep loss and three found no impact of sleep loss [49, 50, 52]. A similar pattern is observed for Pe latency, with one study demonstrating a delay in latency following SR [47], and two reporting no impact of TSD [49, 52].

Post-error behaviour and correction

Seven studies examined post-error adjustments [33-35, 45, 48, 49, 52]. Two reported increased error rate, and reduced correct rate in post-error trials following sleep loss [33, 52]. Murphy et al. [45] found reduced post-error slowing after extended wake. However, Lust [35] and Kusztor et al. [49] reported no significant effect of sleep loss on post-error adjustments or post-error slowing. Interestingly, Hsieh et al. [34] found the influence of sleep loss on post-error adjustments depended upon whether the errors were corrected. Trials following corrected errors were more likely to be correct, less likely to be missed, and were faster, compared to uncorrected errors, and this relationship was particularly evident during TSD. Two studies examined error correction: one found TSD increased correction time and decreased correction rate [34], whereas Hsieh et al. [48] found TSD did not influence correction time, and only influenced correction rate for incongruent trials.

Discussion

The first aim of this review was to summarise the existing literature investigating the impact of sleep loss on performance monitoring and error-monitoring.

Performance monitoring

Limited consensus regarding the impact of sleep loss on performance monitoring accuracy was observed, with no discernible differences in findings between TSD and SR studies. Of interest, where sleep loss did significantly influence performance monitoring accuracy (58% of accuracy studies)[7-9, 12, 35, 37, 38], the result for the majority of tasks (9 of 14) was more conservative estimates of performance, such that overestimations of performance became more accurate (two tasks), underestimations of performance worsened (three tasks), accurate estimations became underestimations (three tasks), or accurate estimations became more accurate (one task). This pattern would suggest participants can appreciate decreases in performance, even if they are unable to accurately assess their exact performance level. Alternatively, more conservative estimates of performance may instead reflect an *expectation* of poor performance following sleep loss, rather than an actual recognition of performance deficits. The meta-analysis also identified a pattern of underestimation of performance, although in contrast to the systematic review, found underestimations were reduced during sleep loss. However, it is important to note no studies demonstrating overestimations of performance following sleep loss were included in the meta-analysis as these studies did not report the adequate information required for inclusion. As a result, these data should be interpreted with caution.

Similar to performance monitoring accuracy, there was limited consensus regarding the impact of sleep loss on the relationship between objective performance and subjective ratings of performance. This is likely due to significant methodological differences in testing conditions, sleep manipulation protocols, cognitive tasks employed, and statistical approach. However, in all studies, even those without significant correlations between objective performance and subjective performance ratings, subjective ratings did change in the expected direction following sleep loss [39-41, 54]. This again suggests participants may recognise decreases in performance, but this recognition is not detectable using a correlation. These findings may also reflect an expectation of poorer performance following sleep deprivation.

Overall, calibration was not influenced by sleep loss, but it was task-dependent [14]. Two studies [12, 43] reported reduced confidence-accuracy correlations during TSD, however this was limited to seemingly random testing sessions within individual studies. Therefore, our overall conclusion from this set of studies is the more confident individuals feel in the accuracy of their response, the more likely they are to be correct, and this does not change during sleep deprivation. While this suggests individuals have insight into their performance, we emphasise high confidence does not necessarily mean good performance. Specifically, confidence-accuracy correlations can only tell us the relative magnitude of the relationship between these two variables and not the accuracy of performance monitoring. Additionally, investigators have recently advised against using confidence-accuracy correlations in measuring metacognition [10], due to the inability to control for the influence of bias (whether one is predisposed to high or low confidence), which affects confidence-accuracy relationships [10]. Given the heterogeneity in methods and findings across performance monitoring studies, it is unclear whether certain tasks, tasks measuring specific cognitive domains, or certain features of tasks (e.g., difficulty level) confer more risk for inaccurate self-monitoring. Boardman et al. [9] suggest task difficulty influences performance monitoring accuracy, such that underestimation of performance was greater on a more difficult version of a task. Conversely, Baranksi et al.[11] found participants demonstrate overconfidence on more difficult tasks, and that confidence is less likely to distinguish between correct and incorrect responses when a task is considered very easy or particularly difficult. Through standardisation of methods, and the specific investigation of difficulty levels with a task, future research can better ascertain whether certain tasks are more at risk for inaccurate performance monitoring and miscalibration.

Error-monitoring

Within the error monitoring literature, we examined studies investigating the impact of sleep loss on error detection, error awareness, and/ or post-error adjustments. Findings within this body of literature were more consistent, largely reporting a reduction in ERN and Pe amplitude following sleep loss, suggesting sleep loss impairs both error-detection and error-awareness. This finding was supported by the results of the meta-analysis and remained even when conservatively removing Hsieh et al [33, 48] to account for the influence of error correction and monetary incentives on error-monitoring. Of note, although the ERN and Pe are considered reliable measures of error-detection and error-awareness, no included studies in this review utilised behavioural measures to examine changes in trial by trial error awareness, as has been conducted outside of the sleep field [20, 21, 25, 55]. Pe amplitude is specific to trials on which participants are consciously aware they have made mistakes [25]. By not calculating Pe amplitude separately for “aware” and “unaware” trials, Pe amplitude results are potentially diminished by the inclusion of trials where the Pe is either minimal or absent due to a lack of error-awareness. Therefore, the results reported here may be an underestimation of the real effect. Furthermore, although Pe and ERN amplitude provide some indication of whether the error-monitoring system is intact, behavioural measures of error-awareness allow researchers to specifically examine whether sleep loss reduces the proportion of errors which make it into conscious awareness. Error correction is also important to consider when discussing error-awareness, given logically, we must first recognise an error to correct it. Only two studies [34, 48] within this review measured error-correction and both found it was reduced following sleep loss. This reduction in correction rate provides further evidence of error-awareness impairment following sleep loss. Of interest, Hsieh et al., [48] did not find a significant difference in Pe amplitude between well-rested and sleep deprived conditions despite a significant difference in correction rate, further suggesting the use of the Pe alone to measure error-awareness is limited.

Very few studies investigated the impact of sleep loss on Pe and ERN latency and post-error adjustments, with mixed results. Interpretations of this literature are therefore difficult, and further research is required to make definitive conclusions. Behavioural latency of error awareness as well as the latency of the Pe are particularly important for future research to consider. Recent research suggested latency of the error awareness behavioural response (i.e., the individual signalling they made an error) predicts the latency of the emergence of the Pe [25]. Therefore, studies which do not measure the timing of error awareness, may not be measuring Pe at the timepoint when it is most likely to occur. In addition, a decrease in average Pe amplitude can be due to an increase in the variability of the Pe onset across trials rather than a decrease in Pe amplitude at the single-trial level. This, in turn, suggests that the reported reduction in Pe amplitude could be explained by an increase in temporal variability rather than a genuine decrease in the amplitude of the Pe signal. Thus, significant differences in Pe amplitude between well-rested and sleep deprived conditions may potentially be driven by differences in the timing of awareness, rather than impairments in the system itself [25].

Methodological differences, similarities, and quality

The second aim of this study was to identify key methodological similarities and differences among studies. Key similarities within the performance monitoring literature include the use of within-subjects designs and the age of participants (largely young adults aged 18-25). The main methodological differences in this literature, included the method utilised to measure performance-monitoring (subjective estimates, vs subjective self-ratings vs confidence) and the type of cognitive task assessed. Differences in performance monitoring method appear to be the largest factor contributing to inconclusive results across studies. We provide recommendations for standardising this literature when discussing future directions below. In contrast to the performance monitoring literature, the error-monitoring literature is distinctly more consistent in methodology. Key similarities include: a) participants age (mostly between 18 and 25 years old); b) the type of task used, with almost all studies employing an inhibitory control task; c) the use of TSD of ~ 24 hours; and d) the electrode sites used to measure ERPs of interest (ERN at FCz, and Pe at Cz.) No significant differences in methodology were observed among studies.

The third aim of this review was to examine the methodological quality of the literature to determine the risk of bias, and to examine whether studies employed methods to appropriately address research questions. Common strengths of both performance monitoring and error monitoring studies included low risk of bias in sample selection and, overall, low risk of confounds. Most studies utilising within subjects' designs were counterbalanced and included multiple measures of performance-monitoring pre and post the sleep manipulation, allowing effects of sleep loss to be distinguished from effects of repeated measures. This was less common in the error-monitoring literature, so future studies may want to consider increasing the number of observations within studies to explore the plausibility of alternative explanations of variations in performance. Additionally, it was common for both performance monitoring and error-monitoring studies to provide inadequate information about assumptions, normality, and whether data was modified to account for any violations. This made it difficult to determine whether statistical analyses were appropriate, and future research must be mindful to provide detailed information on how data was handled prior to analysis. Within the performance monitoring literature, all studies used methods which addressed a relevant aspect of performance monitoring. However, studies should be mindful of the limitations of their chosen measures, acknowledging only what they can truly reveal about the relationship between sleep and performance monitoring and be cautious not to over interpret findings. This is particularly true for the studies utilising correlational designs, where investigators often interpreted a significant correlation as evidence of awareness of performance, or no correlation as evidence of an inability to monitor performance. Furthermore, studies often interpret a change in correlation strength or significance between well-rested and sleep deprived conditions as a difference between conditions despite no statistical test examining whether conditions are significantly different.

Limitations and future directions

The final aim of this review was to identify limitations of the current literature and identify important next steps. The performance monitoring literature is limited by the methods used to measure performance monitoring. Metacognitive sensitivity – the extent to which confidence discriminates between correct and incorrect trials – can be confounded by metacognitive bias - the extent to which an individual is predisposed to report high or low confidence on a given task [10]. Additionally, metacognitive sensitivity is influenced by task performance, whereby an individual will better distinguish correct from incorrect responses on a task they deem easy [56]. This is important, as no studies included in this review controlled for bias or task performance, therefore the true effect of sleep loss on sensitivity remains unclear. Performance monitoring studies utilising just one estimate or rating of subjective performance either before or after the task are also likely influenced by bias. This is even more important within a sleep loss context as individuals within this environment may expect their performance to worsen following sleep deprivation. Thus, if one is predisposed to low confidence or conservative estimates of performance, this bias may be further amplified during sleep loss. Consequently, discrepancies between objective and subjective performance under sleep loss observed here may instead reflect bias or changes in bias, rather than the inability of an individual to accurately monitor performance. Controlling for the influence of bias and task performance are therefore critical next steps in understanding the true effect of sleep loss on performance monitoring. To achieve this, future studies must: 1) examine trial by trial estimates of accuracy or confidence, and; 2) consider robust statistical measures such as meta-*d*', which control for response bias and task performance confounds [56]. Future research should also consider including metacognitive training prior to experimental sessions, as evidence suggests metacognitive ability does improve with repeated exposure [57].

The main limitation of the error-monitoring literature is a reliance on the Pe alone to examine the impact of sleep loss on error-awareness, without behavioural indication of error awareness. Reductions in Pe amplitude after sleep loss are suggested to reflect impairments in error-awareness, however none of the included studies obtained behavioural data to determine whether participants consciously detected errors or not. Thus, whether sleep loss is a risk factor for errors to go unnoticed remains unclear. To address this limitation, future studies should employ both behavioural and EEG measures to investigate whether reductions in Pe amplitude indeed reflect reduced error-awareness. However, it should be noted, requiring participants to signal errors as they occur may have counteracting effects which reduce the impact of sleep loss on error-awareness. Specifically, the requirement to signal an error as it occurs may encourage participants to attend more carefully to their performance, and thus make them more likely to detect errors when they occur. Future research must take this into consideration when interpreting findings. Further investigations on post-error adjustments will also inform this literature, as theories suggest error correction may be an indication an individual consciously recognised an error [15]. To incorporate measures of post-error adjustments investigators must consider conducting analyses to examine post-error trials for accuracy and RT changes, and utilise tasks allowing participants to correct errors.

For both performance monitoring and error-monitoring, relatively few studies investigate chronic SR. As individuals are far more likely to be sleep restricted than completely sleep deprived, future research should investigate the impact of multiple days of SR to gain an accurate understanding of how both performance monitoring and error monitoring manifest under real world conditions. More SR research will also allow future studies to determine whether performance monitoring and error-monitoring are differentially affected by TSD and SR, which was not possible in the current review, given the limited number of SR studies. In addition, most studies did not control for sleep duration or quality prior to conducting sleep loss protocols, therefore reported results may be confounded by differences in pre-study conditions. To be sure any effects on performance or error-monitoring are not confounded by pre-study sleep habits, future studies should control for and monitor sleep and wake patterns prior to experimental sessions.

In conclusion, the impact of sleep loss on performance monitoring is dependent upon the method used to measure performance monitoring. Although the accuracy of self-assessed performance varied under sleep loss conditions, on tasks where sleep loss significantly influenced participants' ability to accurately monitor their performance, participants demonstrated more conservative estimates of performance. Confidence-accuracy relationships did not appear to be influenced by sleep loss and no consensus could be reached regarding the impact of sleep loss on objective performance and subjective accuracy ratings, due to inconsistent methods. The error-monitoring literature suggests sleep loss may impair both unconscious error-detection and conscious error-awareness processes. This review highlighted two important methodological limitations of current research: 1) the influence of bias within metacognitive measures, and; 2) the lack of behavioural methods in error-monitoring studies to substantiate changes observed in EEG measures. Therefore, future research must employ improved methodological and statistical measures. Performance monitoring research should leverage the methods developed by psychophysicists in the framework of Signal Detection Theory and implement bias-free measures of metacognition such as the meta-*d'*. Error monitoring research should collect behavioural reports about participants' awareness of their errors in parallel with EEG. These additions will enable more definitive conclusions to be made on how performance monitoring and error-monitoring are influenced by sleep loss. Such research could identify the neural mechanisms decreasing or preserving performance and error monitoring and has implications in the development of fatigue management strategies which may reduce the dangerous consequences of sleep loss.

References

1. Anderson C, Ftouni S, Ronda JM, Rajaratnam SMW, Czeisler CA, Lockley SW. Self-reported Drowsiness and Safety Outcomes While Driving After an Extended Duration Work Shift in Trainee Physicians. Sleep. 2017;41(2).
2. Lee ML, Howard ME, Horrey WJ, Liang Y, Anderson C, Shreeve MS, et al. High risk of near-crash driving events following night-shift work. PNAS. 2016;113(1):176.
3. Tefft BC. Prevalence of motor vehicle crashes involving drowsy drivers, United States, 1999–2008. Accid Anal Prev. 2012;45:180-6.
4. Adams RJ, Appleton SL, Taylor AW, Gill TK, Lang C, McEvoy RD, et al. Sleep health of Australian adults in 2016: results of the 2016 Sleep Health Foundation national survey. Sleep Health. 2017;3(1):35-42.
5. Faraut B, Andrillon T, Vecchierini MF, Leger D. Napping: A public health issue. From epidemiological to laboratory studies. Sleep Med Rev. 2017;35:85-100.
6. Rouault M, McWilliams A, Allen MG, Fleming SM. Human Metacognition Across Domains: Insights from Individual Differences and Neuroimaging. Personal Neurosci. 2018;1:e17.
7. Baranski JV, Gil V, McLellan TM, Moroz D, Buguet A, Radomski M. Effects of modafinil on cognitive performance during 40 hr of sleep deprivation in a warm environment. Mil Psychol. 2002;14(1):23-47.
8. Baranski JV, Pigeau RA. Self-monitoring cognitive performance during sleep deprivation: effects of modafinil, d-amphetamine and placebo. J Sleep Res. 1997;6(2):84-91.
9. *Boardman JM, Bei B, Mellor A, Anderson C, Sletten TL, Drummond SPA. The ability to self-monitor cognitive performance during 60 h total sleep deprivation and following 2 nights recovery sleep. J Sleep Res. 2018;27(4):e12633.
10. *Fleming SM, Lau HC. How to measure metacognition. Front Hum Neurosci. 2014;8(443).
11. Baranski JV, Pigeau RA, Angus RG. On the ability to self-monitor cognitive performance during sleep deprivation: a calibration study. J Sleep Res. 1994;3(1):36-44.

12. Blagrove M, Akehurst L. Effects of sleep loss on confidence-accuracy relationships for reasoning and eyewitness memory. J Exp Psychol Appl. 2000;6(1):59-73.

13. Harrison Y, Horne JA. Sleep loss and temporal memory. Q J Exp Psychol [A]. 2000;53(1):271-9.

14. Baranski JV. Fatigue, sleep loss, and confidence in judgment. J Exp Psychol Appl. 2007;13(4):182-96.

15. Rabbitt PM. Errors and error correction in choice-response tasks. J Exp Psychol Appl. 1966;71(2):264-72.

16. Danielmeier C, Ullsperger M. Post-error adjustments. Front Psychol. 2011;2:233.

17. Luck SJ. An Introduction to the Event-Related Potential Technique. Cambridge, UNITED STATES: MIT Press; 2014.

18. *Gehring WJ, Goss B, Coles MGH, Meyer DE, Donchin E. A Neural System for Error Detection and Compensation. Psychol Sci. 1993;4(6):385-90.

19. Charles L, Van Opstal F, Marti S, Dehaene S. Distinct brain mechanisms for conscious versus subliminal error detection. Neuroimage. 2013;73:80-94.

20. Endrass T, Reuter B, Kathmann N. ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. Eur J Neurosci. 2007;26(6):1714-20.

21. *Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GPH, Kok A. Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. Psychophysiology. 2001;38(5):752-60.

22. Yeung N, Summerfield C. Metacognition in human decision-making: confidence and error monitoring. Philos Trans R Soc Lond B Biol Sci. 2012;367(1594):1310-21.

23. Hester R, Foxe JJ, Molholm S, Shpaner M, Garavan H. Neural mechanisms involved in error processing: A comparison of errors made with and without awareness. NeuroImage. 2005;27(3):602-8.

24. Coles MGH, Scheffers MK, Holroyd CB. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. Biol Psychol. 2001;56(3):173-89.

25. *Murphy PR, Robertson IH, Allen D, Hester R, O'Connell RG. An electrophysiological signal that precisely tracks the emergence of error awareness. Front Hum Neurosci. 2012;6:65.

26. Yeung N, Botvinick MM, Cohen JD. The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. Psychol Rev. 2004;111(4):931-59.

27. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. PLoS Med. 2009;6(7):e1000100.

28. Innovation VH. Covidence systematic review software Melbourne, Australia [Available from: www.covidence.org.

29. Tufanaru C, Munn Z, Aromataris E, J. C, Hopp L. Chapter 3: Systematic reviews of effectiveness. 2020. In: JBI Manual for Evidence Synthesis [Internet]. JBI. Available from: https://synthesismanual.jbi.global.

30. Mitchell M, Muftakhidinov B, Winchen T. Engauge Digitizer Software.

31. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539-58.

32. Lin L, Chu H. Quantifying publication bias in meta-analysis. Biometrics. 2018;74(3):785-94.

33. Hsieh S, Li TH, Tsai LL. Impact of monetary incentives on cognitive performance and error monitoring following sleep deprivation. Sleep. 2010;33(4):499-507.

34. *Hsieh S, Tsai CY, Tsai LL. Error correction maintains post-error adjustments after one night of total sleep deprivation. J Sleep Res. 2009;18(2):159-66.

35. Lust SA. To sleep, perchance to decide: The effect of sleep deprivation on error reactivity [dissertation on the Internet]. Columbia: University of Missouri; 2013.

36. Steward JM, Pilcher JJ, Scheck-Bradley P. Effects of sleep deprivation on actual and estimated performance on complex cognitive tasks. Sleep 2009. p. A139-A40.

37. Terlizzese T, Hsiou D, Zeter D, Chapagain N, Gao C, Ginty A, et al. Experimental sleep restriction, anticipatory stress, and learning organic chemistry. Sleep 2019. p. A41-A2.

38. Daviaux Y, Mignardot J-B, Cornu C, Deschamps T. Effects of total sleep deprivation on the perception of action capabilities. Exp Brain Res. 2014;232(7):2243-53.

39. Biggs SN, Smith A, Dorrian J, Reid K, Dawson D, van den Heuvel C, et al. Perception of simulated driving performance after sleep restriction and caffeine. J Psychosom Res. 2007;63(6):573-7.

40. Kosmadopoulos A, Sargent C, Zhou X, Darwent D, Matthews RW, Dawson D, et al. The efficacy of objective and subjective predictors of driving performance during sleep restriction and circadian misalignment. Accid Anal Prev. 2017;99(Pt B):445-51.

41. Philip P, Sagaspe P, Taillard J, Moore N, Guilleminault C, Sanchez-Ortuno M, et al. Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. Sleep. 2003;26(3):277-80.

42. Smith BP, Browne M, Armstrong TA, Ferguson SA. The accuracy of subjective measures for assessing fatigue related decrements in multi-stressor environments. Saf Sci. 2016;86:238-44.

43. Aidman E, Jackson SA, Kleitman S. Effects of sleep deprivation on executive functioning, cognitive abilities, metacognitive confidence, and decision making. Appl Cogn Psychol. 2019;33(2):188-200.

44. Mathew GM, Strayer SM, Ness K, Chang AM. Chronic sleep restriction increases confidence in incorrect responses during a working memory task. Sleep Med. 2019;64 (Supplement 1):S247.

45. Murphy TI, Richard M, Masaki H, Segalowitz SJ. The effect of sleepiness on performance monitoring: I know what I am doing, but do I care? J Sleep Res. 2006;15(1):15-21.

46. Asaoka S, Fukuda K, Murphy TI, Abe T, Inoue Y. The effects of a nighttime nap on the error-monitoring functions during extended wakefulness. Sleep. 2012;35(6):871-8.

47. Gossard TR, Westerland SM, Linn-Evans M, Timm PC, Sandness DJ, Dueffert L, et al. The cumulative effect of partial chronic sleep restriction on the neural processing stream in neurologically normal individuals. Sleep. 2019;42.

48. Hsieh S, Cheng IC, Tsai LL. Immediate error correction process following sleep deprivation. J Sleep Res. 2007;16(2):137-47.

49. *Kusztor A, Raud L, Juel BE, Nilsen AS, Storm JF, Huster RJ. Sleep deprivation differentially affects subcomponents of cognitive control. Sleep. 2019;42(4):01.

50. Renn RP, Cote KA. Performance monitoring following total sleep deprivation: effects of task type and error rate. Int J Psychophysiol. 2013;88(1):64-73.

51. Scheffers MK, Humphrey DG, Stanny RR, Kramer AF, Coles MG. Error-related processing during a period of extended wakefulness. Psychophysiology. 1999;36(2):149-57.

52. Tsai LL, Young HY, Hsieh S, Lee CS. Impairment of error monitoring following sleep deprivation. Sleep. 2005;28(6):707-13.

53. Liang Y, Ko C, Tsai L, Hsieh S. Effects of one night of total sleep deprivation on error processing. Sleep 2012. p. A112.

54. Dorrian J, Lamond N, Dawson D. The ability to self-monitor performance when fatigued. J Sleep Res. 2000;9(2):137-44.

55. *Hester R, Foxe JJ, Molholm S, Shpaner M, Garavan H. Neural mechanisms involved in error processing: a comparison of errors made with and without awareness. Neuroimage. 2005;27(3):602-8.

56. *Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious Cogn. 2012;21(1):422-30.

57. Schwiedrzik CM, Singer W, Melloni L. Subjective and objective learning effects dissociate in space and in time. PNAS. 2011;108(11):4506.

Practice Points

1. Performance monitoring literature addresses three main questions related to: 1) performance monitoring accuracy; 2) subjective ratings and objective performance relationships; and, 3) confidence-accuracy relationships.
2. Within performance monitoring accuracy studies, results are inconsistent, although overall findings suggest sleep loss leads to more conservative estimates of performance.
3. Within subjective ratings data, methodologies are inconsistent, therefore no definitive conclusions can be made on whether sleep loss affects the relationship between accuracy and subjective ratings.
4. As confidence increases, so does the proportion of correct responses, and sleep loss does not impact this relationship.
5. Sleep loss is associated with amplitude reductions in the error-related negativity and error-positivity, suggesting unconscious and conscious error detection are impaired following sleep deprivation and sleep restriction.

Research Agenda

1. Where possible, future performance monitoring studies should obtain trial by trial estimates of confidence or performance and conduct statistical analyses which account for metacognitive bias and task performance (e.g meta-*d*').
2. Where methodological or operational constraints prevent the collection of trial by trial data, studies should obtain global estimates of performance accuracy which can be directly compared to objective performance outcomes and avoid the use of subjective rating scales.

4. The error-monitoring literature must incorporate behavioural measures of conscious error recognition to confirm changes in error-positivity are reflective of impaired error awareness.
5. Further research examining latency of the error-positivity will determine if changes observed between well-rested and sleep deprived conditions are truly reflective of impairments in error-monitoring, or actually reflect the latency of error awareness responses.

Table 1. The impact of sleep loss on performance monitoring outcomes

Author (year)	**Task/s**	**Impact of sleep loss on accuracy**		**Confidence-accuracy relationship**	**Objective performance-subjective performance relationship**
		Pre-task/trial	Post-task/trial		
Performance monitoring literature					
Aidman et al., 2019 [43]	Medical decision-making			Sig. corr ↓	
Baranski et al. 1994 [11]	Serial addition			Sig. corr × Well-calibrated ×	
Baranski & Pigeau 1997 [8]	Perceptual Comparison Mental Addition	Overconfident ✓ Accurate ×	Overconfident ✓ Accurate ×		
Baranski et al. 2002 [7]	Logical Reasoning Vigilance Addition Comparison	Underestimate ✓ Overestimate✓ Accurate× Accurate×	Underestimate ✓ Overestimate✓ Accurate× Accurate×		
Baranski 2007 [14]	General Knowledge Perceptual Comparison Mental Addition	Overestimate× Accurate× Accurate×	Overestimate× Accurate× Underestimate×	Mis-calibrated × Well-calibrated× Well-calibrated×	
Biggs et al. 2007 [39]	York Driving Simulator				NS correlation ×

Author (year)	Task/s	Impact of sleep loss on accuracy		Confidence-accuracy relationship	Objective performance-subjective performance relationship
		Pre-task/trial	Post-task/trial		
Blagrove & Akehurst 2000 [12]	Logical Reasoning		Underestimate✓		
	Raven's matrices		Accurate✓		
	Suggestibility			Sig. neg corr✓	
Boardman et al. 2018 [9]	Serial Subtraction of 1		Accurate✓		
	Serial Subtraction of 3		Accurate×		
	Serial Subtraction of 7		Underestimate✓		
	PVT		Accurate✓		
Daviaux et al. 2014 [38]	Stepping over capability	Underestimate✓			
Harrison & Horne 2000 [13]	Temporal Memory			TSD more confident on incorrect trials ✓	
Hsieh et al. 2009 [34]	Flanker		Underestimate×		
Hsieh et al. 2010 [33]	Flanker		Underestimate×		
Kosmadopolus et al. 2017 [40]	York Driving Simulator				Sig. corr↓
Lust 2013 [35]	Weapons Identification Task		Overestimate✓		
Mathew et al. 2019 [44]	Visual object learning task			TSD more confident on incorrect trials ✓	

Author (year)	Task/s	Impact of sleep loss on accuracy		Confidence-accuracy relationship	Objective performance-subjective performance relationship
		Pre-task/trial	Post-task/trial		
Philip et al., 2003 [41]	Reaction time task				NS corr ↓
Smith et al., 2016 [42]	PVT	↑ Discrepancy between objective/subjective performance✓			
Steward et al. 2009 [36]	GRE		Overestimate×		
Terlizzese et al., 2019 [37]	Chemistry exam	Underestimate✓ (females only)			

Note: DAT = dental admissions test, GRE = graduate record exam LSAT = law school admissions test, N/A = not applicable, PVT = psychomotor vigilance task.

Note: each results column first describes performance during sleep loss (accurate, underestimation, overestimation of accuracy; sig = significant effect, NS = not significant) and then describes how performance changed with sleep loss (× =no effect of sleep loss, ✓= significant effect of sleep loss; ↓=reduction or ↑=increase in correlation). Study [8] found accurate self-monitoring of perceptual comparison up to 54 hours of total sleep deprivation, after which participants were overconfident.

Table 2. Impact of sleep loss on error-monitoring.

Error-monitoring literature							
Author (year)	**Task**	**Effect of sleep loss on Amplitude**		**Effect of sleep loss on Latency**		**Effect of sleep loss on post-error behaviour**	
Asaoka et al. 2012 [46]	Flanker task	ERN: ×	Pe: ×				
Gossard et al. 2019 [47]	Visual attention task	ERN: ↓	Pe: ↓	ERN: N/A	Pe: ↑		
Hsieh et al. 2007 [48]	Flanker	ERN: ↓	Pe: ×			Correction %↓	Correction ms×
Hsieh et al. 2009 [34]	Flanker					Correction % ↓ Post error-accuracy ×	Correction ms ↑
Hsieh et al. 2010 [33]	Flanker	ERN: ↓	Pe: ↓	ERN: ↑	Pe: N/A	Post-error accuracy ↓	
Kusztor et al. 2019 [49]	Stop Signal Task	ERN: ×	Pe: ↓	ERN: ×	Pe: ×		Post-error slowing×
Liang et al. 2012 [53]	Flanker	EPP: ↓					
Lust 2013 [35]	Weapons Identification Task	ERN: ×	Pe: ×			Post-error behaviour ×	
Murphy et al. 2006 [45]	Flanker Task	ERN: ×	Pe: ↓				Post-error slowing↓
Renn & Cote 2013 [50]	Flanker Task	ERN: ×	Pe: ↑				
	GoNogo Task	ERN: ↓	Pe: ×	ERN: ×	Pe: N/A		
Scheffers et al. 1999 [51]	Visual Task	ERN: ↓	Pe: N/A				
	Memory Task	ERN: ↓	Pe: N/A				

Tsai et al. 2005 [52]	Flanker Task	ERN: ↓	Pe: ↓	ERN: ×	Pe: ×	Post-error accuracy ↓

Note: ERN – error-related negativity, N/A = not applicable, Pe- error positivity, ↓=reduction, ↑=increase, × =no effect of sleep loss, ✓= significant effect of sleep loss, sig = significant effect. Study [34] - decrease in ERN only evident in group receiving incentives. Study [35] - post-error accuracy evident only in corrected errors. Study [48] – correction % and ERN amplitude only reduced for incongruent trials. Study [50] - larger Pe was observed in sleep loss group when results were collapsed across block.

Supp Table 2. Characteristics of included studies

First author (year)	Publication type	N (f)	Within or Between	Sleep manipulation (length)	Cognitive task/s	Pre vs Post	Performance monitoring/error-monitoring method
Aidman et al. 2019 [43]	Peer-reviewed paper	13 (0)	Within	TSD (42 hours)	Decision-making	Post-trial	Confidence/accuracy
Asaoka et al. 2012 [46]	Peer-reviewed paper	20 (7)	Between	TSD (22 hours)	Arrow orientation task	N/A	ERN & Pe Amplitude
Baranski et al. 1994 [11]	Peer-reviewed paper	16 (0)	Within	TSD (46 hours)	Serial addition	Post-trial	Calibration curve Confidence/accuracy
Baranski & Pigeau 1997 [8]	Peer-reviewed paper	41 (2)	Within	TSD (64 hours)	1. Perceptual Comparison 2. Mental Addition	Pre & Post task	Subjective estimates
Baranski et al. 2002 [7]	Peer-reviewed paper	5 (0)	Within	TSD (40 hours)	1. Mental Addition 2. Detection of repeated numbers 3. Logical Reasoning 4. Perceptual comparison	Pre & post task	Subjective estimates
Baranski 2007 [14]	Peer-reviewed paper	64 (25)	Within	TSD (28 hours)	1. General Knowledge 2. Mental Addition 3. Perceptual comparison	Pre & post-task	Subjective estimates

First author (year)	Publication type	N (f)	Within or Between	Sleep manipulation (length)	Cognitive task/s	Pre vs Post	Performance monitoring/error-monitoring method
Biggs et al. 2007 [39]	Peer-reviewed paper	12 (6)	Within	SR (1N 4 hours)	1. York Driving Simulator	Pre-task	1. Subjective performance scale
Blagrove et al. 2000 [12]	Peer-reviewed paper	93 (45)	Between	TSD (50 hours)	2. Logical reasoning 3. Abstract reasoning 4. Gudhonsson Suggestibility scale	Post-task	2. Metacognition 3. Subjective estimate
Boardman et al. 2018 [9]	Peer-reviewed paper	40 (22)	Within	TSD (60 hours)	1. PVT 2. Mental addition	Post-task	1. Subjective scale (transformed) 2. Subjective estimate
Daviaux et al. 2014 [38]	Peer-reviewed paper	20 (12)	Between	TSD (24 hours)	Stepping over capability	Pre-task	Subjective estimate
Gossard et al. 2019 [47]	Peer-reviewed paper	18 (8)	Within	SR (9 nights 4 hours)	Visual attention task	N/A	ERN & Pe Amplitude Pe Latency
Harrison & Horne 2000 [13]	Peer-reviewed paper	40 (20)	Mixed	TSD (35 hours)	1. Temporal memory 2. Self-ordered pointing	Post-trial	Metacognition
Hsieh et al. 2007 [48]	Peer-reviewed paper	16 (7)	Within	TSD (24 hours)	Flanker task	N/A	ERN and Pe Amplitude Correction rate
Hsieh et al. 2009 [34]	Peer-reviewed paper	16 (7)	Within	TSD (24 hours)	Flanker task	Post-task	Subjective performance estimate ERN Latency Post error adjustment
Hsieh et al. 2010 [33]	Peer-reviewed paper	24 (12)	Within	TSD (24 hours)	Flanker task	Post-task	Subjective performance estimate

First author (year)	Publication type	N (f)	Within or Between	Sleep manipulation (length)	Cognitive task/s	Pre vs Post	Performance monitoring/error-monitoring method
							ERN & Pe Amplitude
Kosmadopolus et al. 2017 [40]	Peer-reviewed paper	32 (0)	Between	SR (4 hours)	York Driving Task	Pre-task	Subjective performance scale
Kusztor et al. 2019 [49]	Peer-reviewed paper	24 (12)	Within	TSD (24 hours)	Stop signal task	N/A	ERN & Pe Amplitude ERN & Pe Latency Post-error adjustments
Liang et al. 2012 [53]	Abstract	9 (3)	Within	TSD (24 hours)	Flanker	N/A	Error-preceding positivity
Lust et al. 2013 [35]	Dissertation	42 (25)	Between	SR (4 hours)	Weapons Identification Task	Post-trial	Confidence & Accuracy ERN & Pe Amplitude ERN & Pe Latency
Mathew et al. 2019 [44]	Abstract	15 (0)	Within	SR (5 nights 5 hours)	Visual object learning task	Post-trial	Confidence & Accuracy
Murphy et al. 2006 [45]	Peer-reviewed paper	17 (17)	Within	TSD (20 hours)	Flanker	N/A	ERN & Pe Amplitude Post-error adjustments
Philip et al. 2003 [41]	Peer-reviewed paper	10 (0)	Within	SR (2 hours)	Reaction time task	Pre-task	Subjective performance scale

First author (year)	Publication type	N (f)	Within or Between	Sleep manipulation (length)	Cognitive task/s	Pre vs Post	Performance monitoring/error-monitoring method
Renn & Cote 2013 [50]	Peer-reviewed paper	49 (25)	Between	TSD (34 hours)	1. Flanker 2. GoNogo	N/A	ERN & Pe Amplitude ERN & Pe Latency
Scheffers et al. 1999 [51]	Peer-reviewed paper	8 (0)	Within	TSD (24 hours)	1. Memory search task 2. Visual search task	N/A	ERN Amplitude
Smith et al. 2016 [42]	Peer-reviewed paper	91 (12)	Between	SR (4 hours)	1. PVT	Pre-task	Subjective scale
Steward et al. 2009 [36]	Abstract	23 (8)	Within	TSD (24 hours)	2. Graduate Record Exam 3. Dental Admissions Test	Post-task	Subjective estimate
Terlizzese et al. 2019 [37]	Abstract	100 (64)	Between	SR (6 hours)	Organic chemistry exam	Pre-task	Subjective estimate
Tsai et al. 2005 [52]	Peer-reviewed paper	16 (7)	Within	TSD (24 hours)	Flanker task	N/A	ERN and Pe Amplitude ERN & Pe Latency Post-error adjustments

Note. ERN = error-related negativity, N/A = not applicable, Pe = error-positivity, SR = sleep restriction, TSD = total sleep deprivation, Pre = subjective rating/estimate obtained prior to task, Post = subjective rating/estimate obtained post task or trial, PVT = psychomotor vigilance task.