1 Data attributes

All attributes of the working data "bt_with_history_hour_encoded.RData" are listed below:

- subj: Subject ID.
- game_played: Task.
- kss: KSS score.
- falsealarm: Total number of false alarms after performing the task.
- light.task.log: The log-transformed photopic illuminance measured through the phone.
- logMEDI_mean: The log of the mean of the past 24 hours MEDI observations without the 0.1 observations.
- logMEDI_median: The log of the median of the past 24 hours MEDI observations without the 0.1 observations.
- logMEDI_sd: The log of the standard deviation of the past 24 hours MEDI observations without the 0.1 observations.
- logMEDI_min: The log of the minimum of the past 24 hours MEDI observations without the 0.1 observations.
- logMEDI_max: The log of the maximum of the past 24 hours MEDI observations without the 0.1 observations.
- logMEDI_total: The log of the sum of the past 24 hours MEDI observations without the 0.1 observations.
- logMEDI_mean_hour_1_before to logMEDI_mean_hour_24_before : The hourly mean of the log-transformed MEDI observations in each hour in the past 24-hour. For example, logMEDI_mean_hour_6_before is the mean of the MEDI observations from the past 5 hours to the past 6 hours.
- logMEDI_amplitude: Amplitude of the cosinor fit on the 24 hour MEDI history.
- logMEDI_mesor: MESOR of the cosinor fit on the 24 hour MEDI history.

- logMEDI_acrophase: Acrophase of the cosinor fit on the 24 hour MEDI history.
- TPR: The proportion of correct taps out of all correct responses. This provides a measure of sensitivity. Ranges from 0 to 100.
- FDR: The proportion of false positives among all the correct identifications. Ranges from 0 to 100.
- ACC: The proportion of correctly classified events (both true positives and true negatives) out of all events. This measure is available for all tasks, assessing overall performance correctness. Ranges from 0 to 100.
- meanrt: The average reaction time across all responses. This measure is recorded for all tasks, giving insight into the general speed of responses.
- medianrt: The median reaction time across all responses.
- slow10: The average of the slowest 10% of reaction times, indicating moments of inattention or cognitive delays. This measure is available across all tasks.
- fast10: The average of the fastest 10% of reaction times. This provides a measure of peak cognitive performance. This is calculated for all tasks.
- ies: Calculated by dividing the mean reaction time by the proportion of correct answers, combining speed and accuracy into a single measure.
- FPR: Used in N-back and Visual Search tasks, false positive rate is the proportion of false positives out of all actual negatives, assessing the rate of incorrect positive identifications. Ranges from 0 to 100.
- FOR: Used in N-back and Visual Search tasks. It represents the proportion of false negatives among all actual positives, indicating the likelihood of missing true events. Ranges from 0 to 100.
- lapse: Only exists for PVT. Lapse is defined as the total number of slow reaction times (typically greater than 500 ms), indicating moments of inattention or fatigue. Ranges from 0 to the total number of trials.
- DPRIME: Only exists for N-back. D' is a measure from signal detection theory that indicates the sensitivity of the subject in distinguishing between signal and noise assessing their working memory. Also known as D-prime.

- Slope: Specific to Visual search. Slope represents the rate of change in performance over time or across difficulty levels in the visual search task.
- date: The date when the task was performed.
- hour: The hour of day when the task was performed.
- datetime: The timestamp when the cognitive task session was initiated.
- dayhour: The time of day in hours when the task was initiated.
- game_datetime: The timestamp when the task was initiated.
- timeawake: The time awake in hours.
- count: The total count of the number of observations of the task each subject has.

For more information on the attributes of the questionnaires in the data 'bt_baseline', see the supplementary materials of Didikoglu et al.[1].

2 Imputation

The missing values within the working dataset were dealt with specifically:

- Remove records with missing values from newly added variables;
- Keep the missing value related to task-specific variables such as lapse only available to PVT, similarly for D-prime to N-back, and slope to Visual Search;
- Keep missing values that are unable to calculate such as the false positive rate for PVT when this task has no skip option hence no false negative;
- Remove records with missing true positive rate (TPR) and any reaction-related result.
- Fill correct results based on other available results, such as false omission rate (FOR) must be 0 when TPR is 100.

[1] A. Didikoglu, N. Mohammadian, S. Johnson, and R. J. Lucas, "Associations between light exposure and sleep timing and sleepiness while awake in a sample of UK adults in everyday life," *Proceedings of the National Academy of Sciences*, vol. 120, no. 42, e2301608120, 2023. Available: `https://doi.org/10.1073/pnas.2301608120`

- Records with missing KSS were filled with KSS scores from the other task under the same session.
- Remove visual search observations with missing slope.