

RESEARCH ARTICLE

Open Access



Test-retest reliability of brain arousal regulation as assessed with VIGALL 2.0

Jue Huang^{1*}, Christian Sander^{1,2,3}, Philippe Jawinski^{1,2,3}, Christine Ulke^{2,3}, Janek Spada^{1,3}, Ulrich Hegerl^{1,2,3} and Tilman Hensch^{1,2}

* Correspondence: jue.huang@medizin.uni-leipzig.de
¹Department of Psychiatry and Psychotherapy, University of Leipzig, Semmelweisstr. 10, Leipzig 04103, Germany
Full list of author information is available at the end of the article

Abstract

Background: Different levels of brain arousal can be delineated not only during sleep but also during wakefulness. Electroencephalography (EEG) is the gold standard to assess different levels of brain arousal. A novel EEG- and electrooculography (EOG)-based tool, the Vigilance Algorithm Leipzig (VIGALL 2.0), allows determining the level of EEG-vigilance (indicating brain arousal). Considering the frequency patterns and LORETA-based cortical distribution of electroencephalic activity, VIGALL 2.0 automatically attributes one out of seven vigilance stages to each EEG segment (1-sec EEG segments by default), ranging from high alertness (stage 0), to relaxed wakefulness (stage A1 to A3), to drowsiness (stage B1 to B2/3) up to sleep onset (stage C). Building on the time series of these seven vigilance stages across 20 min, two parameterizations of the temporal dynamic (brain arousal regulation) are calculated: the lability score and the slope index.

Methods: 27 healthy participants (age = 22.93 ± 3.44 years, 18 females) underwent two sessions (7 days apart) of a twenty-minute eyes-closed resting EEG paradigm.

Results: The test-retest reliability coefficients for the EEG-vigilance stages were between $\rho = .53$ and $.86$ (all $p < .01$). For the temporal dynamic of the stages across 20 min, the test-retest reliability coefficients were $\rho = .70$ (lability score, $p < .001$) and $.71$ (slope index, $p < .001$).

Conclusions: This study demonstrated some trait aspects of brain arousal regulation by confirming the stability of temporal dynamic of EEG-vigilance stages as assessed with VIGALL 2.0. Considering the "first day in lab" effect identified in the present study, more adaptation to the lab surrounding and a stricter control of other state factors should be taken into account, which might improve reliability. Additionally, in a clinical context, a broader range of brain arousal regulation patterns might be found, possibly leading to higher test-retest reliability than was found in this homogenous healthy sample. This would be desirable, as parameters of brain arousal regulation are promising diagnostic and prognostic biomarkers for diseases with arousal disturbances, such as affective disorders, ADHD and fatigue.

Keywords: Test-retest reliability, Brain arousal regulation, EEG-vigilance stages, VIGALL 2.0

Background

Brain arousal influences all human behaviours (Pfaff et al. 2008). Arousal regulation denotes the adaptation of brain arousal to situational requirements, which is of most importance for efficient human behavior. For example, arousal must be heightened in case of potential danger, maintained during cognitive tasks and reduced at bedtime. Recently, the Research Domain Criteria (RDoC) project of the National Institute of Mental Health has implemented arousal as a fundamental dimension of psychopathology (Cuthbert and Insel 2013).

Different levels of brain arousal can be delineated not only during sleep (e.g. Iber et al. 2008; Rechtschaffen and Kales 1969) but also during wakefulness (see below). Electroencephalography (EEG) is the gold standard to assess different levels of brain arousal. By taking into account the frequency patterns as well as the Low Resolution Electromagnetic Tomography (LORETA)-based cortical distribution of EEG activity, different EEG-vigilance stages (indicating brain arousal states) can be distinguished during the transition from high alertness to relaxed wakefulness, to drowsiness up to sleep onset.

Under eyes-closed resting conditions without external stimulation, interindividual differences concerning arousal regulation occur (Hegerl and Hensch 2014): Most subjects show progressive declines to lower EEG-vigilance stages indicating *slowly-declining arousal regulation*, while some subjects display an *unstable arousal regulation* with rapid declines to low EEG-vigilance stages after only a few seconds. Yet others exhibit a *hyperstable arousal regulation* without such declines to lower EEG-vigilance stages even after recording periods of 20 min - the EEG recording period following the standard operating procedures of VIGALL (available at <http://research.uni-leipzig.de/vigall/>) and the Multiple Sleep Latency Test (MSLT; Carskadon and Dement 1977). Brain arousal regulation can be considered a state modulated trait, because sleep deficits or drugs with arousal stabilizing properties (e.g. caffeine, nicotine) are likely to exert state dependent influences.

EEG parameters of brain arousal regulation (i.e. the temporal dynamic of EEG-vigilance stages) can be considered converging biomarkers reflecting influences of various wakefulness and sleep promoting neurochemical systems (e.g. the noradrenergic, histaminergic, serotonergic and orexinergic system). These biomarkers are not only of interest as neurobiological correlates of arousal regulation but also as indicators of pathogenetic aspects in affective disorders and attention-deficit/hyperactivity disorder (ADHD) (Geissler et al. 2014; Hegerl and Hensch 2014).

According to the classification by Rechtschaffen and Kales (1969) and the scoring manual of the American Academy of Sleep Medicine (Iber et al. 2008), brain arousal has traditionally been classified into 3 stages: relaxed wakefulness, non-rapid eye movement sleep (NREM) and rapid eye movement sleep (REM). However, these traditional classifications do not distinguish sub-stages *preceding* sleep onset, although such sub-stages during transition from wakefulness to sleep onset have been described for decades (Bente 1964; Loomis et al. 1937; Roth 1961; Benca et al. 1999; Cantero et al. 2002; Corsi-Cabrera et al. 2000; De Gennaro et al. 2001, 2004, 2005; De Gennaro and Ferrara 2003; Kaida et al. 2006; Marzano et al. 2007; Strijkstra et al. 2003; Tsuno et al. 2002). Building on this research, a novel EEG-based algorithm for automatic classification of EEG-vigilance stages has recently been introduced by Hegerl and colleagues (Sander et al. in press).

The Vigilance Algorithm Leipzig (VIGALL) 2.0 is an EEG- and electrooculography (EOG)-based algorithm which allows to objectively determine the level of EEG-vigilance and its dynamics within EEG recordings (Hegerl et al. 2014; Hegerl and Hensch 2014; Olbrich et al. 2012; Sander et al. in press). The algorithm considers the frequency patterns as well as the cortical distribution of EEG activity, using EEG source localisation approaches (LORETA; Pascual-Marqui et al. 1994; Pascual-Marqui et al. 2002). VIGALL 2.0 automatically attributes one out of seven EEG-vigilance stages (Table 1) to each EEG segment (1-sec EEG segments by default). Building on the time series of these seven EEG-vigilance stages across 20 min, two parameterizations of the temporal dynamic are calculated: the lability score and the slope index (as described below).

Because of the high inter-individual variability in electroencephalic frequency patterns and power, VIGALL 2.0 has adaptive features concerning individual alpha peaks and amplitude levels. VIGALL 2.0 should not be used for EEGs showing alpha variant rhythms or major modifications due to drugs (e.g. anticholinergic drugs) or certain diseases (e.g. severe Alzheimer's disease). Additionally, VIGALL 2.0 is not indicated for EEGs from children younger than ten years (or older children in case of delayed maturation). VIGALL 2.0 is implemented as an add-in in the Brain Vision Analyzer 2.0 software (Brain Products, Gilching, Germany). The VIGALL 2.0 add-in as well as the user manual including the standard operating procedures (SOPs) can freely be downloaded from <http://research.uni-leipzig.de/vigall/>.

VIGALL 2.0 is a refinement of an earlier version of the algorithm, which has been validated using simultaneous EEG-fMRI (Olbrich et al. 2009) as well as simultaneous EEG-FDG-PET-studies (Günther et al. 2011) and by relating the EEG-vigilance stages to different autonomic parameters (Olbrich et al. 2011b). Moreover, VIGALL has already been applied in clinical studies. In line with the arousal regulation theory of affective disorders and ADHD (Hegerl and Hensch 2014; Geissler et al. 2014), a hyperstable arousal regulation has been shown in depressive patients compared with controls (Hegerl et al. 2012; Olbrich et al. 2012), and, in contrast, an unstable arousal regulation has been shown in ADHD (Olbrich et al. 2013; Sander et al. 2010).

Having in mind the potential use of EEG parameters of brain arousal regulation as diagnostic or predictive biomarkers, the aim of this study was to assess the test-retest reliability of lability score, slope index, EEG-vigilance stages and mean EEG-vigilance level as assessed by VIGALL in healthy subjects.

Table 1 Arousal states and EEG descriptions of EEG-vigilance stages

VIGALL	arousal state	EEG
0	high alertness	low amplitude EEG with high beta (12–25 Hz) power without horizontal slow eye movements
A1	relaxed wakefulness	dominant high alpha (8–12 Hz) in occipital regions of interest
A2		dominant high alpha (8–12 Hz) in parietal or temporal regions of interest
A3		dominant high alpha (8–12 Hz) in frontal regions of interest
B1	drowsiness	low amplitude EEG with high beta (12–25 Hz) power with horizontal slow eye movements
B2/3		high delta (2–4 Hz) or theta (4–8 Hz) power
C	sleep onset	K-complexes or sleep spindles

Vigilance Algorithm Leipzig (VIGALL)

Methods

Participants

Participants were recruited via local and online advertisements. They gave written informed consent and received an expense allowance. The study was approved by the local ethics committee of the University of Leipzig (075-13-11032013). Participants were requested to participate in two EEG measures with the average interval of 7.15 (± 1.29 ; range: 4–11) days between both. 27 participants (18 females; age = 22.93 ± 3.44 years) had two EEG recordings (T1 and T2) available and met the following inclusion criteria: no pathological sleepiness (defined as score above 14 of Epworth Sleepiness Scale; see Johns 1992), which may artificially increase retest reliability, no extensive sleep duration difference between both nights prior to each session (i.e. more than 2 hours as assessed by sleep protocol) and no reported history of psychiatric or neurological disorders or current consumption of psychotropic medication. Participants were allowed to drink caffeinated beverages and to smoke as usual prior to the EEG recording at T1 and T2.

EEG recording

During EEG preparation participants were asked to fill in questionnaires including a sleep protocol, which contained questions about last night's sleep, a substance consumption scale, the Epworth Sleepiness Scale (ESS, German version available at http://www.charite.de/dgsm/dgsm/fachinformationen_frageboegen_epworth.php) and the Karolinska Sleepiness Scale (KSS; Akerstedt and Gillberg 1990). EEG was recorded within a timeslot between 1 and 4 p.m. depending on time arrangement with the participants. Daytime of EEG assessment was kept constant for each participant across both sessions. The recording booth was a light dimmed and sound attenuated room with a maintained temperature beneath 25 °C. The participants lay comfortably on a lounger with closed eyes and were instructed to relax and not to try staying awake in case of drowsiness. Rather, they were explicitly allowed to follow their natural course of arousal. EEG assessment started with a Berger Maneuver (i.e. alternating opening and closing of the eyes). The resting EEG recording in each session lasted for 20 min.

The EEG was recorded by 40 channel QuickAmp amplifiers (Brain Products GmbH, Gilching, Germany) from 32 electrode sites and no online filter was applied. Electrodes were arranged according to the extended international 10–20 system using EasyCap (EASYCAP Brain Products GmbH, Gilching, Germany), and referenced against common average. Impedances were kept below 10 k Ω and sampling rate was 1000 Hz. Electrodes for horizontal EOG (hEOG) were placed lateral of the left and right eye, electrodes for vertical EOG (vEOG) were placed above and beneath the right eye.

EEG preprocessing and vigilance classification

EEG data were processed using Brain Vision Analyzer 2.0 software (Brain Products GmbH, Gilching, Germany). EEG raw data was filtered offline (highpass at .5 Hz, lowpass at 70 Hz, notch-filter at 50 Hz (± 5 Hz)). EOG raw data was filtered the same way as EEG raw data but without a highpass filter. The 20 min resting EEG was subdivided into 1200 consecutive 1-sec segments. Following a visual artifact screening, an independent component analysis (ICA) was performed. Eye movement and continuous muscle artifacts were removed by

extracting independent components clearly containing artifact-related information (Olbrich et al. 2011a). Afterwards, the EEG segments were visually screened again for remaining muscle, swallow, eye movement and sweating artifacts. To keep the time sequence intact, artifact-containing segments were not removed but marked manually and discarded from further classification. EEG recordings with more than 90 % of artifacts free segments were included in the analysis. This criterion was achieved by all subjects: There were 1173.30 (± 25.53 , range: 1087–1199) and 1172.22 (± 19.00 , range: 1127–1196) artifact free EEG segments at T1 and T2, respectively.

Subsequently, all segments were classified into seven different vigilance stages using VIGALL 2.0 (for detailed descriptions of the scoring algorithm see the VIGALL 2.0 Manual (available at <http://research.uni-leipzig.de/vigall/>). Classification of vigilance stage C relies on the occurrence of graph elements indicating sleep onset (i.e. K-complex or sleep spindles). Therefore, all EEGs had visually been screened for such graph elements and the respective EEG segments had been marked.

EEG-vigilance parameterizations

For each recording, results of the VIGALL classification via VIGALL 2.0 are written into an output text file which was then imported into a customized Excel template with Visual Basic for Applications (VBA) macros (Microsoft), and used to calculate different parameterizations of brain arousal and its dynamics.

- a) *Amount of EEG-vigilance stages*: Over the whole recording period, the relative amount of segments attributed to the respective EEG-vigilance stages (stage 0, A, B and C, sub-stages A1, A2, A3, B1, B2/3) was determined ($\text{amount} \cdot 100 / \text{total number of non-artifact segments}$).
- b) *Mean EEG-vigilance level*: As a parameterization of the overall EEG-vigilance level, each EEG-vigilance stage was assigned with a number ranging from 7 (highest vigilance stage 0) to 1 (lowest vigilance stage C). Based on that, a mean EEG-vigilance level was calculated by averaging the vigilance scores of all non-artifact segments. In order to assess the temporal dynamics of the EEG-vigilance level, the 20 min resting EEG was subdivided into four consecutive blocks (block 1: minute 1–5; block 2: minute 6–10; block 3: minute 11–15; block 4: minute 16–20) and the mean EEG-vigilance level was computed for each block.
- c) *Lability score*: To quantify the speed and extend of vigilance decline, a so called “lability score” was determined for each individual vigilance time course (see Table 2). To this end, epochs of 1 min duration (i.e. 60 1-sec segments) were analyzed (epoch 1: segments 1–60, epoch 2: segments 2-61...). For each epoch, it was tested if one of the following conditions applied (in ascending order according to lability): (I) occurrence of at least 1 C-stage; (II) at least 1/3 of segments classified as B2/3-stage; (III) at least 1/3 of all segments classified as B-stages (B1 + B2/3); (IV) at least 2/3 of all segments classified as 0/A- or 0/A1-stages. If within the whole EEG data only criterion IV was fulfilled, a lability score of 1 or 2 was given. If one of the other criteria was fulfilled in at least one epoch, we determined the EEG part in which the lowest vigilance stage was reached for the first time and gave the respective “lability score” (as seen in Table 2).

Table 2 Lability score correspond to certain EEG-vigilance stage in the respective EEG block (20 min EEG recording separated in 4 blocks of 5 min duration)

Lability Score	Lability level (criterion)	EEG block	Operational definition
1	level 1: less than 1/3 of all segments not classified as 0/A- or 0/A1-stages	1-4	rigidity, unique appearance of 0 and A1
2		1-4	rigidity, unique appearance of 0 and A
3	level 2: at least 1/3 of all segments classified as B (B1 + B2/3)-stages	4	stage B emerged in minute 16-20
4		3	stage B emerged in minute 11-15
5		2	stage B emerged in minute 6-10
6		1	stage B emerged in minute 1-5
7	Level 3: at least 1/3 of segments classified as B2/3-stages	4	stage B2/3 emerged in minute 16-20
8		3	stage B2/3 emerged in minute 11-15
9		2	stage B2/3 emerged in minute 6-10
10		1	stage B2/3 emerged in minute 1-5
11	Level 4: occurrence of at least 1 C-stage	4	stage C emerged in minute 16-20
12		3	stage C emerged in minute 11-15
13		2	stage C emerged in minute 6-10
14		1	stage C emerged in minute 1-5

d) *Slope Index*: In order to measure the degree of arousal instability, we set up functional equations $f(x) = m\sqrt{\ln x + n}$, which take into account the initial mean EEG-vigilance level observed in minute one (n , with $1 \leq n \leq 7$) and its square rooted logarithmic decline to the mean level in each of the subsequent minutes two to twenty ($m\sqrt{\ln x}$ with x representing the count of the respective minute; $2 \leq x \leq 20$). Arousal instability was defined as the slope (m) with the lowest value (reflecting the steepest decline) as derived from the nineteen possible functional equations. The square rooted logarithmic relation enables an objective and adequate weighting of immediate drops to moderate EEG-vigilance levels relative to delayed drops to low EEG-vigilance levels. Within an independent sample comprising 1079 healthy participants of the Leipzig Health Care Study (556 males, age = 65.6 years), regression analyses revealed a linear negative relation between mean EEG-vigilance level during minute one and m ($b = -.33$, $t = -10.875$, $p < .001$). To compensate floor effects within participants exhibiting low initial EEG-vigilance levels, individual slopes were adjusted by adding a linear correction factor $(-0.33 * (7-n))$.

Statistics

Paired two samples Wilcoxon signed ranks tests were carried out to determine whether scores of T1 and T2 differed significantly from each other. For all correlational analyses nonparametric Spearman rank correlation rho was calculated, since the data were either ordinal scaled (lability score and slope index) or not normally distributed (amount of vigilance stages with exception of stage A1 and B2/3).

Results

Test-retest reliability of EEG-vigilance stages

EEG-vigilance stages (in % of all artifact free 1-sec EEG segments) at T1 and T2 are presented in Table 3. Except higher amounts of stage 0 ($Z = -2.30$, $p = .022$) and lower

Table 3 EEG-vigilance stages (in % of all artifact free 1-sec segments across 20 min of EEG recording). Mean differences between T1 and T2 and respective Spearman correlation coefficients (N = 27)

EEG-vigilance stages	T1			T2			Z (p)		rho (p)	
	proportion	(SD)	range	proportion	(SD)	range				
stage 0	5.61	(8.81)	.00-33.73	3.04	(5.63)	.00-23.99	-2.30	(.022)	.529	(.005)
stage A	49.23	(25.83)	1.51-94.43	44.23	(31.54)	2.17-97.46	-1.97	(.049)	.855	(.000)
sub-stage A1	32.08	(17.02)	.50-66.97	30.07	(23.80)	.42-96.10	-1.07	(.285)	.680	(.000)
sub-stage A2	12.56	(14.70)	.17-48.12	10.48	(14.70)	.17-66.16	-1.38	(.167)	.700	(.000)
sub-stage A3	4.59	(8.54)	.00-39.10	3.68	(8.08)	.00-38.43	-1.09	(.276)	.776	(.000)
stage B	36.52	(18.69)	3.51-76.45	40.76	(23.42)	2.20-83.69	-1.39	(.163)	.569	(.002)
sub-stage B1	18.62	(15.28)	.67-50.88	17.15	(16.37)	.34-54.49	-1.09	(.313)	.590	(.001)
sub-stage B23	17.89	(14.12)	.09-48.28	23.61	(17.30)	.00-57.18	-2.15	(.032)	.657	(.000)
stage C	8.65	(12.18)	.00-41.05	11.97	(17.23)	.00-56.06	-1.24	(.215)	.859	(.000)

amounts of B2/3 at T1 compared with T2 ($Z = -2.15, p = .032$), EEG-vigilance stages did not show any statistically significant differences between T1 and T2. Notice that the significant difference was also obtained for entire A ($Z = -1.97, p = .049$) but not for the comparisons between sub-stage A1, A2 and A3 (all $p > .05$). The test-retest correlation coefficients were all significant (all $p < .01$; see Table 3).

Test-retest reliability of mean EEG-vigilance level

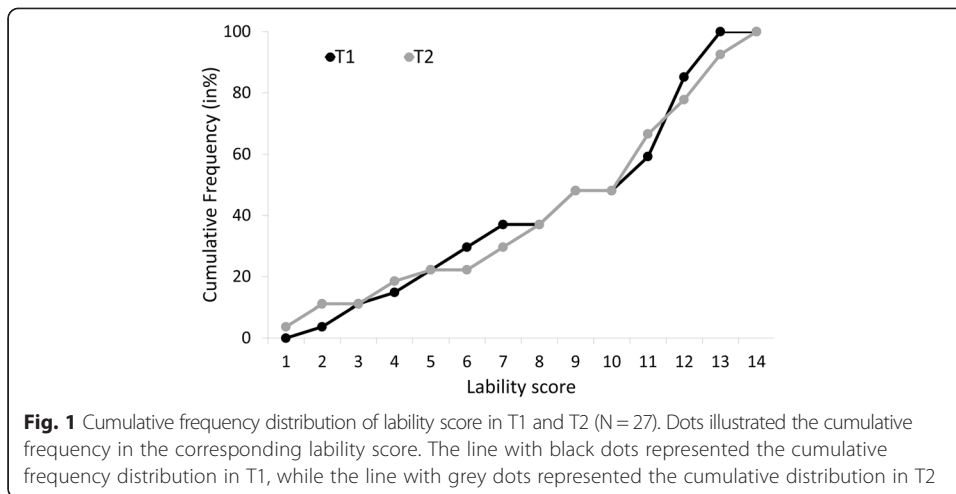
Mean EEG-vigilance level for 20 min (overall) and the 5–minutes blocks are shown in Table 4. The overall EEG-vigilance level was significantly lower in T2 ($Z = -2.35, p = .019$). A significant difference was also found for mean EEG-vigilance level of the first block (minute 1–5) ($Z = -2.22, p = .026$). The effect size between T1 and T2, as presented in Table 4, for overall and mean EEG-vigilance level in corresponding time blocks varied between .19 and .49, which indicated a small to medium effect based on Cohen’s criterion (Cohen 1988). The correlation coefficients are also illustrated in Table 4. With exception of the first block (minute 1–5), all mean vigilance levels demonstrated temporal stability with correlation coefficients of $\rho \geq .659$ (all $p < .001$; see Table 4).

Test-retest reliability of temporal dynamic: lability score

The cumulative frequency distribution and scatter plot of the lability scores in T1 and T2 are illustrated respective in Figs. 1 and 2. The test-retest correlation of lability score was $\rho = .704 (p < .001)$.

Table 4 Mean EEG-vigilance level for 20 min (overall) and the 5–minutes blocks. Differences between T1 and T2 and respective Spearman correlation coefficients (N = 27)

Time block	T1 (SD)		T2 (SD)		effect size d	Z (p)		rho (p)	
overall	4.13	(0.95)	3.80	(1.17)	.31	-2.35	(.019)	0.700	(.000)
1	5.26	(0.66)	4.83	(1.05)	.49	-2.22	(.026)	0.377	(.052)
2	4.35	(1.09)	4.01	(1.37)	.27	-1.71	(.088)	0.758	(.000)
3	3.68	(1.31)	3.39	(1.43)	.21	-1.44	(.149)	0.692	(.000)
4	3.23	(1.49)	2.95	(1.46)	.19	-1.23	(.220)	0.659	(.000)

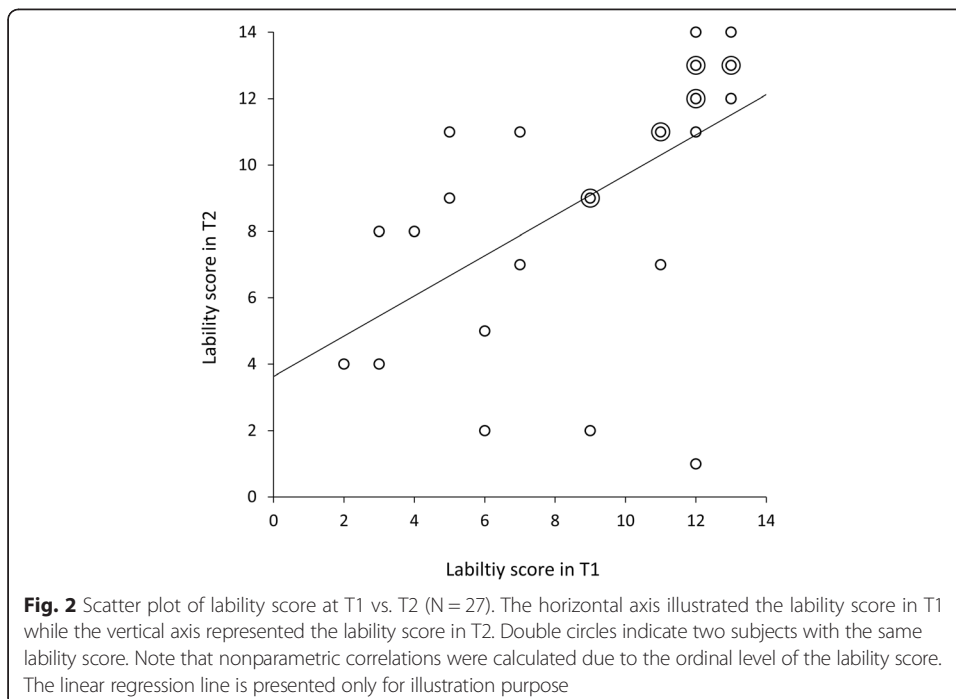


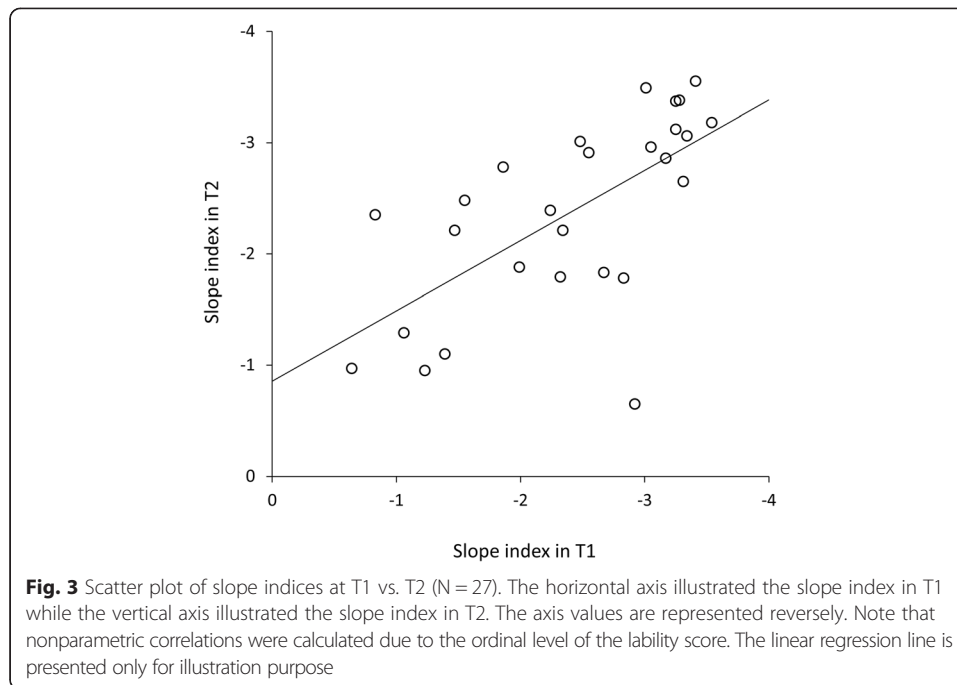
Test-retest reliability of temporal dynamic: slope index

The slope indices for T1 and T2 are represented as scatter plot in Fig. 3. The test-retest correlation of logarithmic slope index to the corrected sample size was $\rho = .714$ ($p < .001$).

Discussion

The current study analyzed for the first time test-retest reliability of EEG parameters of brain arousal regulation in a sample of young individuals. The nonparametric test-retest correlation coefficients for single EEG-vigilance stages varied from .53 to .86 (all $p < .01$), those for mean EEG-vigilance levels with exception of block 1 between .66 and .76 (all $p < .001$), and





those for parameterization of the temporal dynamic (lability score and slope index) were .70 and .71, respectively (all $p < .001$). Thus, the EEG parameters, as classified with VIGALL 2.0, turned out to be temporally stable, confirming some trait aspects of brain arousal regulation.

The obtained reliabilities in the present study are well suited for applications on group level, but should further be improved for clinical purposes, for example by more strictly considering influencing factors as discussed below. In this context it should also be noted that we analyzed a homogenous healthy sample and excluded subjects with excessive sleepiness. Given this reduction of variance, high reliabilities are difficult to obtain. One might speculate that in a clinical context, a broader range of brain arousal regulation patterns will be found, possibly leading to higher test-retest reliabilities than found in this homogenous healthy sample. This hypothesis is based on the findings in previous studies that a pronounced instability or hyperstability of arousal has been found in mania/ADHD and depression, respectively. This disease-related variance should stay constant, whereas in healthy subject, in contrast, a floor effect, or a generally reduced variance might reduce correlations.

The parameterization of mean EEG-vigilance levels (in the current study for 20 min and for 5-min blocks) proved to be reliable. On this basis, entering consecutive blocks of vigilance into repeated measurement analyses enables to reliably assess group-dependent differences in the temporal dynamics of mean EEG-vigilance levels. Such an approach has, for instance, been pursued by Jawinski and colleagues comparing mean EEG-vigilance levels among participants carrying zero, one or two susceptibility alleles, respectively, of certain genetic polymorphisms linked to Bipolar Disorder (Jawinski et al. unpublished). In our study the subjects showed significantly higher amount of stage 0 ($Z = -2.30$, $p = .022$) and stage A ($Z = -1.97$, $p = .049$) and lower amount of B2/3 ($Z = -2.15$, $p = .032$) at T1 compared to T2, indicating overall higher arousal at T1. In

line with these results, mean EEG-vigilance level in the first 5-min block at T1 was significantly higher ($Z = -2.22$, $p = .026$) than that at T2, and the EEG-vigilance level in the following 5-min blocks showed a tendency in the same direction, again indicating higher arousal at T1 (effect sizes between .19 to .49). These findings might be a consequence of some “first day in lab” effects, i.e. an increased arousal due to the unfamiliar surroundings or to anticipatory anxiety. In sleep medicine such a laboratory-induced effect resulting in impaired sleep quality is known as “first night effect” (Agnew et al. 1966; Tamaki et al. 2005). Consequently, the validity of the parameters provided by VIGALL 2.0 is supported, however, these findings suggest that such laboratory-induced factors should be controlled more strictly. Longer adaptation to the lab, repeated assessments or defining more strictly a baseline arousal by a demanding task before starting the quiet rest period might further improve reliability.

Sleep duration can be another severe confounder in vigilance assessment (Rosenthal et al. 1993a, Rosenthal et al. 1993b), especially in samples of young subjects who display large variability in night sleep (Lo et al. 2014). In the current study, information about sleep was derived by self-reports and no information about validity of this assessment is available. In future studies objective assessments of sleep-wake-behavior or quality of sleep prior to the EEG should be included, which could be achieved using actigraphy or polysomnography. In addition, in the current study the subjects had not been asked about recent stressors (e.g. examinations), which are considered as an important reason for disturbed sleep or arousal (Sanford et al. 2014; Winzeler et al. 2014). Considering such state effects more rigorously could further improve reliability. Another possible confounder is the consumption of arousal modulating drugs, such as caffeine and nicotine (Jaehne et al. 2012; Johnson et al. 1990). This might be especially the case in clinical populations, where not only smoking is highly prevalent (Lasser et al. 2000), but also several arousal-impairing medications come into play (e.g. antidepressants; Hensch et al. 2015). In the current healthy sample, however, smoking and caffeine was ad libitum and no subject was excluded, as no excessive consume was reported and as we tried to avoid withdrawal effect. Moreover, we did not control for effects of the menstrual phase. If we had assessed our female subjects in the same menstrual phase at T1 and T2, we might have increased the reliability. Instead, we tried to keep our subjects as much as possible in a “real life” condition which might have resulted in a more conservative estimation of the effect.

The MSLT (Carskadon and Dement 1977) was so far the only established method to assess vigilance regulation. The MSLT, however, only assesses EEG-defined sleep onset, disregarding information about fluctuations of vigilance before sleep onset. The MSLT has a complex testing protocol, which requires four or five 20-min trials with 2 hours between each trial and strongly tries to control for several of the mentioned confounders: Normally, the subjects arrive at the laboratory on the evening before the MSLT. Thus, their MSLT testing is done on the second day, which probably reduces the “first day in lab” effects. Additionally, the MSLT protocol controls factors such as sleep duration in the preceding night and tobacco use within 30 min before testing. Furthermore, before testing only light meals are recommended and caffeine and bright sunlight should be avoided (for more details see Littner et al. 2005). Comparing the test-retest reliabilities obtained in the current study with those for the MSLT is difficult due to the dearth of studies. Almost all studies on retest reliability are on clinical groups where

reliability varies probably in accordance with varying symptomatology: Retest reliability was non-significant in patients with hypersomnia of different etiology over a mean test-retest interval of 4.2 years ($\rho = .17$; Trotti et al. 2013), low in insomniacs over eight months ($\rho = .44$; Roehrs et al. 2011), and good in narcolepsy over an interval of 5–21 days ($r = .81$; Folkerts et al. 2008). Only one study in healthy subjects is available (Zwyghuizen-Doorenbos et al. 1988), which reported a high test-retest reliability of MSLT in 14 participants ($r = .97$, $p < .001$). A limitation of this study is certainly the small sample size in combination with parametric correlations, which are easily inflated by one or very few subjects. Nonetheless, the study by Zwyghuizen-Doorenbos et al. (1988) is instructive as the authors also calculated the reliabilities when the number of the MSLT test runs was reduced. The reliability was between $.65$ ($p < .01$) and $.79$ ($p < .008$) when the test was repeated for only two times (combined test runs 10 a.m. together with 12 p.m., and 2 p.m. with 4 p.m., respectively). These reliabilities of the combined scores of two MSLT runs are numerically comparable to the test-retest reliability in the present study. When only one single 20 min MSLT run was analyzed, reliability of the MSLT was no longer significant for all test runs which began at afternoon (12 p.m., 2 p.m. and 4 p.m.) (Zwyghuizen-Doorenbos et al. 1988). These findings and the current results suggest that repeated EEG recordings might be important in cases where high reliability is needed, such as clinical applications.

Another limitation of present study is the short test interval. A period of one week does not allow clearly separating trait and state aspects of a variable. Arousal regulation is strongly associated with psychopathological states, such as a depressive or manic episode, which would both last more than one week. However, in the current sample we only analyzed healthy subjects probably limiting the influence of permanent mood states on test retest reliability.

Conclusion

EEG parameters of brain arousal regulation are promising diagnostic and prognostic biomarkers for diseases with arousal disturbances, such as affective disorders, ADHD and fatigue (Geissler et al. 2014; Hegerl et al. (2012); Hegerl and Hensch 2014; Kluge et al. 2013; Olbrich et al. 2012). Here, we demonstrated in healthy subjects that the temporal dynamics of EEG-vigilance stages across 20 min are relatively stable individual characteristics when assessed with VIGALL 2.0. These inter-individual differences are likely to influence results in several cognitive tasks and neurobiological assessments such as fMRI and FDG-PET (Günther et al. 2011; Olbrich et al. 2009). By a more strict control of state factors modulating arousal regulation (e.g. preceding sleep, nicotine, caffeine, adaptation to the study procedure and environment) even higher reliabilities might be achievable.

Conflict of interest

Within the last three years, Prof. Hegerl was an advisory board member for Lilly, Lundbeck, Takeda Pharmaceuticals, Servier and Otsuka Pharma and a speaker for Bristol-Myers Squibb, Medice Arzneimittel, Novartis and Roche Pharma. The other authors do not declare any conflict of interest.

Author's contributions

JH, UH and TH conceived and designed the experiments. JH performed the experiments. JH, CS, PJ, JS and TH analyzed the data. JH, CS, PJ, CU, UH and TH wrote and revised the paper. All authors read and approved the final manuscript.

Acknowledgement

We acknowledge support from the German Research Foundation (DFG) and Universität Leipzig within the program of Open Access Publishing.

Author details

¹Department of Psychiatry and Psychotherapy, University of Leipzig, Semmelweisstr. 10, Leipzig 04103, Germany.

²LIFE – Leipzig Research Center for Civilization Diseases, Universität Leipzig, Leipzig, Germany. ³Depression Research Centre of the German Depression Foundation, Leipzig, Germany.

Received: 31 March 2015 Accepted: 2 June 2015

Published online: 06 October 2015

References

- Agnew HW, Webb WB, Williams RL (1966) The first night effect: an EEG study of sleep. *Psychophysiology* 2(3):263–6
- Akerstedt T, Gillberg M (1990) Subjective and objective sleepiness in the active individual. *Int J Neurosci* 52(1–2):29–37. doi:10.3109/00207459008994241
- Benca RM, Obermeyer WH, Larson CL, Yun B, Dolci I, Kleist KD, Weber SM, Davidson RJ (1999) EEG alpha power and alpha power asymmetry in sleep and wakefulness. *Psychophysiology* 36(4):430–6. doi:10.1111/1469-8986.3640430
- Bente D (1964) Vigilanz, dissoziative Vigilanzverschiebung und Insuffizienz des Vigilanztonus. In: Kranz H, Heinrich K (eds) Begleitwirkungen und Misserfolge der psychiatrischen Pharmakotherapie. Germany, Stuttgart
- Cantero JL, Atienza M, Salas RM (2002) Human alpha oscillations in wakefulness, drowsiness period, and REM sleep: different electroencephalographic phenomena within the alpha band. *Neurophysiol Clin/Clin Neurophysiol* 32(1):54–71. doi:10.1016/S0987-7053(01)00289-1
- Carskadon MA, Dement WC (1977) Sleepiness and sleep state on a 90-min schedule. *Psychophysiology* 14(2):127–33. doi:10.1111/j.1469-8986.1977.tb03362.x
- Cohen J (1988) *Statistical power analysis for the behavioral science*, 2nd edn. Lawrence Erlbaum Associates, Hillsdale. ISBN 978-0805802832
- Corsi-Cabrera M, Guevara MA, Del Río-Portilla Y, Arce C, Villanueva-Hernández Y (2000) EEG bands during wakefulness, slow-wave and paradoxical sleep as a result of principal component analysis in man. *Sleep* 23(6):738–44
- Cuthbert BN, Insel TR (2013) Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 11:126. doi:10.1186/1741-7015-11-126
- De Gennaro L, Ferrara M (2003) Sleep spindles: an overview. *Sleep Med Rev* 7(5):423–40. doi:10.1053/smr.2002.0252
- De Gennaro L, Ferrara M, Curcio G, Cristiani R (2001) Antero-posterior EEG changes during the wakefulness–sleep transition. *Clin Neurophysiol* 112(10):1901–11. doi:10.1016/S1388-2457(01)00649-6
- De Gennaro L, Vecchio F, Ferrara M, Curcio G, Rossini PM, Babiloni C (2004) Changes in fronto-posterior functional coupling at sleep onset in humans. *J Sleep Res* 13(3):209–17. doi:10.1111/j.1365-2869.2004.00406.x
- De Gennaro L, Vecchio F, Ferrara M, Curcio G, Rossini PM, Babiloni C (2005) Antero-posterior functional coupling at sleep onset: changes as a function of increased sleep pressure. *Brain Res Bull* 65(2):133–40. doi:10.1016/j.brainresbull.2004.12.004
- Folkerts M, Rosenthal L, Roehrs T, Krstevska S, Murlidhar A, Zorick F, Wittig R, Roth T (2008) The reliability of the diagnostic features in patients with narcolepsy. *Biol Psychiatry* 40(3):214. doi:10.1016/0006-3223(95)00383-5
- Geissler J, Romanos M, Hegerl U, Hensch T (2014) Hyperactivity and sensation seeking as autoregulatory attempts to stabilize brain arousal in ADHD and mania? *ADHD* 6(3):159–73. doi:10.1007/s12402-014-0144-z
- Günther T, Schönknecht P, Becker G, Olbrich S, Sander C, Hesse S, Meyer PM, Luthardt J, Hegerl U, Sabri O (2011) Impact of EEG-vigilance on brain glucose uptake measured with [(18) F] FDG and PET in patients with depressive episode or mild cognitive impairment. *NeuroImage* 56(1):93–101. doi:10.1016/j.neuroimage.2011.01.059
- Hegerl U, Hensch T (2014) The vigilance regulation model of affective disorders and ADHD. *Neurosci Biobehav Rev* 44:45–57. doi:10.1016/j.neubiorev.2012.10.008
- Hegerl U, Wilk K, Olbrich S, Schönknecht P, Sander C (2012) Hyperstable regulation of vigilance in patients with major depressive disorder. *World J Biol Psychiatry* 13(6):436–46. doi:10.3109/15622975.2011.579164
- Hegerl U, Hensch T, Böttger D, Sander C (2014) VIGALL 2.0: Analyzing different functional brain states and their regulation during resting states. Brain Products Press Release, <http://pressrelease.brainproducts.com/vigall/>
- Hensch T, Blume A, Böttger D, Sander C, Niedermeier N, Hegerl U (2015) Yawning in depression: worth looking into. *Pharmacopsychiatry* 48(3):118–20
- Iber C, Ancoli-Israel S, Chesson A, Quan SF (2008) *Das AASM-Manual zum Scoring von Schlaf und assoziierten Ereignissen: Regeln, Technologie und technische Spezifikationen*, 1st edn. Steinkopff, Germany
- Jaehne A, Unbehauen T, Feige B, Lutz UC, Batra A, Riemann D (2012) How smoking affects sleep: A polysomnographical analysis. *Sleep Med* 13(10):1286–92. doi:10.1016/j.sleep.2012.06.026
- Johns MW (1992) Reliability and factor analysis of the Epworth sleepiness scale. *Sleep* 15(4):376–81
- Johnson LC, Spinweber CL, Gomez SA (1990) Benzodiazepines and caffeine: effect on daytime sleepiness, performance and mood. *Psychopharmacology (Berl)* 101:160–7
- Kaida K, Takahashi M, Åkerstedt T, Nakata A, Otsuka Y, Haratani T, Fukasawa K (2006) Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clin Neuropathol* 117(7):1574–81. doi:10.1016/j.clinph.2006.03.011
- Kluge M, Hegerl U, Sander C, Dietzel J, Mergl R, Bitter I, Demyttenaere K, Gusmão R, Gonzalez-Pinto A, Perez-Sola V, Vieta E, Juckel G, Zimmermann US, Bauer M, Sienaert P, Quintão S, Edel MA, Bolyos C, Ayuso-Mateos JL, López-García P (2013) Methylphenidate in mania project (MEMAP): study protocol of an international randomised double-blind placebo-controlled study on the initial treatment of acute mania with methylphenidate. *BMC Psychiatry* 13:71. doi:10.1186/1471-244x-13-71
- Lasser K, Boyd JW, Woolhandler S, Himmelstein DU, McCormick D, Bor DH (2000) Smoking and mental illness: A population-based prevalence study. *JAMA* 284(20):2606–10. doi:10.1001/jama.284.20.2606

- Littner MR, Kushida C, Wise M, Davila DG, Morgenthaler T, Lee-Chiong T, Hirshkowitz M, Daniel LL, Bailey D, Berry RB, Kapen S, Kramer M, Standards of Practice Committee of the American Academy of Sleep Medicine (2005) Practice parameters for clinical use of the multiple sleep latency test and the maintenance of wakefulness test. *Sleep* 28(1):113–21
- Lo JC, Leong RL, Loh KK, Dijk DJ, Chee MW (2014) Young adults' sleep duration on work days: Differences between East and West. *Front Neurol* 5:81. doi:10.3389/fneur.2014.00081
- Loomis AL, Harvey EN, Hobart GA (1937) Cerebral states during sleep, as studied by human brain potentials. *J Exp Psychol* 21(2):127–44
- Marzano C, Fratello F, Moroni F, Pellicciari MC, Curcio G, Ferrara M, Ferlazzo F, De Gennaro L (2007) Slow eye movements and subjective estimates of sleepiness predict EEG power changes during sleep deprivation. *Sleep* 30(5):610–6
- Olbrich S, Mulert C, Karch S, Trenner M, Leicht G, Pogarell O, Hegerl U (2009) EEG-vigilance and BOLD effect during simultaneous EEG/fMRI measurement. *NeuroImage* 45(2):319–32. doi:10.1016/j.neuroimage.2008.11.014
- Olbrich S, Jödicke J, Sander C, Himmerich H, Hegerl U (2011a) ICA-based muscle artefact correction of EEG data: What is muscle and what is brain? *NeuroImage* 54(1):1–3. doi:10.1016/j.neuroimage.2010.04.256
- Olbrich S, Sander C, Matschinger H, Mergl R, Trenner M, Schönknecht P, Hegerl U (2011b) Brain and Body. *J Psychophysiol* 25(4):190–200. doi:10.1027/0269-8803/a000061
- Olbrich S, Sander C, Minkwitz J, Chittka T, Mergl R, Hegerl U, Himmerich H (2012) EEG vigilance regulation patterns and their discriminative power to separate patients with major depression from healthy controls. *Neuropsychobiology* 65(4):188–94. doi:10.1159/000337000
- Olbrich S, Olbrich H, Jahn I, Sander C, Adamaszek M, Hegerl U, Reque F, Stengler K (2013) EEG-vigilance regulation during the resting state in obsessive-compulsive disorder. *Clin Neuropathol* 124(3):497–502. doi:10.1016/j.clinph.2012.08.018
- Pascual-Marqui RD, Esslen M, Kochi K, Lehmann D (2002) Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Methods Find Exp Clin Pharmacol* 24(Suppl C):91–5
- Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int J Psychophysiol* 18(1):49–65. doi:10.1016/0167-8760(84)90014-X
- Pfaff D, Ribeiro A, Matthews J, Kow LM (2008) Concept and mechanisms of generalized central nervous system arousal. *Ann N Y Acad Sci* 1129:11–25. doi:10.1196/annals.1417.019
- Rechtschaffen A, Kales A, University of California, LABIS (1969) A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects: A. Rechtschaffen and A. Kales (Editors). 26(6): 644–702. doi: 10.1016/0013-4694(69)90021-2.
- Roehrs TA, Randall S, Harris E, Maan R, Roth T (2011) MSLT in primary insomnia: stability and relation to nocturnal sleep. *Sleep* 34(12):1647–52. doi:10.5665/sleep.1426
- Rosenthal L, Roehrs TA, Rosen A, Roth T (1993a) Level of sleepiness and total sleep time following various time in bed conditions. *Sleep* 16(3):226–32
- Rosenthal L, Roehrs TA, Roth T (1993b) The sleep-wake activity inventory: A self-report measure of daytime sleepiness. *Biol Psychiatry* 34(11):810–20. doi:10.1016/0006-3223(93)90070-T
- Roth B (1961) The clinical and theoretical importance of EEG rhythms corresponding to states of lowered vigilance. *Electromyogr Clin Neurophysiol* 13(3):395–9. doi:10.1016/0013-4694(61)90008-6
- Sander C, Arns M, Olbrich S, Hegerl U (2010) EEG-vigilance and response to stimulants in paediatric patients with attention deficit/hyperactivity disorder. *Clin Neurophysiol* 121(9):1511–8. doi:10.1016/j.clinph.2010.03.021
- Sander C, Hensch T, Wittekind DA, Böttger D, Hegerl U (in press) Assessment of wakefulness and vigilance regulation. *Neuropsychobiology*.
- Sanford LD, Suchecki D, Meerlo P (2014) Stress, Arousal, and Sleep, *Current Topics in Behavioral Neurosciences*. doi:10.1007/7854_2014_314
- Strijkstra AM, Beersma DG, Drayer B, Halbesma N, Daan S (2003) Subjective sleepiness correlates negatively with global alpha (8–12 Hz) and positively with central frontal theta (4–8 Hz) frequencies in the human resting awake electroencephalogram. *Neurosci Lett* 340(1):17–20. doi:10.1016/S0304-3940(03)00033-8
- Tamaki M, Nittono H, Hayashi M, Hori T (2005) Examination of the first-night effect during the sleep-onset period. *Sleep* 28(2):195–202
- Trotti LM, Staab BA, Rye DB (2013) Test-retest reliability of the multiple sleep latency test in narcolepsy without cataplexy and idiopathic hypersomnia. *J Clin Sleep Med* 9(8):789–95. doi:10.5664/jcsm.2922
- Tsuno N, Shigeta M, Hyoki K, Kinoshita T, Ushijima S, Faber PL, Lehmann D (2002) Spatial organization of EEG activity from alertness to sleep stage 2 in old and younger subjects. *J Sleep Res* 11(1):43–51. doi:10.1046/j.1365-2869.2002.00288.x
- Winzeler K, Voellmin A, Schäfer V, Meyer AH, Cajochen C, Wilhelm FH, Bader K (2014) Daily stress, presleep arousal, and sleep in healthy young women: a daily life computerized sleep diary and actigraphy study. *Sleep Med* 15(3):359–66. doi:10.1016/j.sleep.2013.09.027
- Zwyghuizen-Doorenbos A, Roehrs T, Schaefer M, Roth T (1988) Test-retest reliability of the MSLT. *Sleep* 11(6):562–5