RELIABILITY OF SLEEP MEASURES

JULIE MOSES, ARDIE LUBIN, PAUL NAITOH AND LAVERNE C. JOHNSON

REPORT NUMBER: 71-45

US NAVY • NEUROPSYCHIATRIC RESEARCH UNIT

NAVY MEDICAL
NEUROPSYCHIATRIC RESEARCH UNIT
SAN DIEGO, CALIFORNIA 92152

BUREAU OF MEDICINE AND SURGERY DEPARTMENT OF THE NAVY
WASHINGTON, D.C. 20390

8

DISCLAIMER NOTICE

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Navy Medical Neuropsychiatric Research Unit San Diego, California 92152	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Reliability of sleep measures

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Julie MOSES, Ardie LUBIN, Paul NAITOH, & Laverne C. JOHNSON

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1971	8	12

8a. CONTRACT OR GRANT NO. b. PROJECT NO. MR005.02 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) 71-45 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Bureau of Medicine and Surgery Department of the Navy Washington, D. C. 20390

13. ABSTRACT

The reliability of sleep measures was calculated over two nights (and within the nights) for twenty young adult males. Percent time in stages 1, 2, 3, and 4, percent movement time, number of movements, and number of stage changes were significantly correlated between subjects over nights. The percent REM time and REM cycle duration were not significantly correlated over nights. Within subjects, the length of the REM period had a significant negative correlation with the length of the preceding NREM period but not with the following NREM period. These data raise questions as to the use of the standard sleep measures as a reliable human traits in young male adults.

DD FORM 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Humans						
Sleep						
Reliability						

UNCLASSIFIED
Security Classification

Psychophysiology
Copyright © 1972 by The Society for Psychophysiological Research

Vol. 9, No. 1
Printed in U.S.A.

Reliability of Sleep Measures

J. Moses, A. Lubin, P. Naitoh, and L. C. Johnson

Navy Medical Neuropsychiatric Research Unit, San Diego

ABSTRACT

The reliability of sleep measures was calculated over two nights (and within the nights) for 20 young adult males. Percent time in stages 1, 2, 3, and 4, percent movement time, number of movements, and number of stage changes were significantly correlated between *Ss* over nights. The percent REM time and REM cycle duration were not significantly correlated over nights. Within *Ss*, the length of the REM period had a significant negative correlation with the length of the preceding NREM period but not with the following NREM period. These data raise questions as to the use of the standard sleep measures as reliable human traits in young male adults.

DESCRIPTORS: Humans, Sleep, Reliability. (J. Moses)

It is a common practice in sleep research to discard the first night in the laboratory because of the first night effects (Agnew, Webb, & Williams, 1966). The general assumption is then made that sleep measures for subsequent nights are consistent from night to night. While many studies have been concerned with the consistency and predictability of sleep characteristics (Williams, Agnew, & Webb, 1964, 1966; Hartmann, 1968; Weitzman, Kripke, Goldmacher, McGregor, & Nogeire, 1970), and report similar group means over several nights, there are only two published studies that report reliability data for individual subjects over nights. Webb and Agnew (1969) computed correlations for sleep stages over 3 nights for nonclinical subjects whose ages ranged from 21 months to 59 years. Johnson, Burdick, and Smith (1970) have reported correlations of sleep characteristics over 12 nights in chronic alcoholics during withdrawal. In both reports, NREM sleep was found to be consistently more reliable than REM sleep.

This research was supported in part by Department of the Navy, Bureau of Medicine and Surgery, under Task No. MR006.02 and by National Science Foundation Grant to the San Diego State College Foundation, No. GB14829. The opinions and assertions contained herein are the private ones of the writers and are not to be construed as official or as reflecting the views of the Navy Department.

Address requests for reprints to: J. Moses, Navy Medical Neuropsychiatric Research Unit, San Diego, California 92152.

This paper reports the analysis of sleep data from 2 nights of uninterrupted sleep in a group of 20 healthy young adults. The purpose was to determine if there are measures of sleep which are reliable and can be used as predictors of other measures of sleep, both within subjects and between subjects.

Method

Twenty male naval recruits, aged 17 to 21, lived in our sleep laboratory for 11 days. They were studied one at a time during a baseline period of 4 nights, and during various experimental periods of sleep deprivation and recovery. Each *S* reported in on Monday and slept that night in the sleep laboratory, attached to all standard electrodes and transducers. These electrodes and transducers were not coupled to our recording equipment and no actual recordings were run. On the 3 subsequent nights, Tuesday, Wednesday, and Thursday, the *S* was recorded. Data from the third and fourth baseline nights (nights 3 and 4) were used in this report.

Ss slept in a sound-treated electrostatically shielded room. Twelve channels of recordings were made on an Offner-Beckman Type R Dynograph. Left and right EOGs were recorded from the outer canthi referenced to the opposite mastoids (A1 and A2). EMG was recorded from the chin. EEG was recorded from the left central (C3), right central (C4), and left occipital portions of the scalp (O1) referenced to the opposite mastoids. Skin potential and skin resistance were recorded using O'Connell-Tursky electrodes on the right hand and forearm. Heart rate was recorded with a Beckman model 9857 cardiotachometer. Finger pulse amplitude and respiration were recorded using mercury strain gauges. Recordings were run at a speed of 10 mm per second.

*S*s were told "go to sleep" at about 2230 and were awakened at about 0600, to provide them with about 7 to 7½ hours of sleep. The range was from 412 to 458 minutes, except for 2 *S*s who only had 390 min due to apparatus difficulties at the beginning of the night.

All records were scored for sleep stages in 30-sec epochs according to the criteria in the Manual of the Association for the Psychophysiological Study of Sleep (Rechtschaffen & Kales, 1968). The percent agreement for epoch by epoch scoring in this laboratory varies from 85 to 94 percent. All the records were scored by the same person, whose scoring reliability is consistently near 90 percent. Any set of successive 30-sec epochs scored as stage REM was called a "REM episode." When REM episodes were separated by less than 15 min, the entire time interval from the onset of the first REM episode to the offset of the last REM episode was called a "REM period." This agrees with the criteria used by previous investigators (e.g., Hartmann, 1968; Globus, 1970).

TABLE 1

Means, standard deviations, and between-night correlations of sleep measures[a]

Sleep Measures	Statistic	Night 3	Night 4	r (N3 × N4)
%Wake	$\bar{X}$	2.065	3.030	.301
	s	1.435	2.573	
%Stage 1	$\bar{X}$	4.265	5.030	.540**
	s	2.233	2.140	
%REM	$\bar{X}$	25.515	23.970	.191
	s	6.002	4.757	
%Stage 2	$\bar{X}$	50.560	51.055	.440*
	s	6.096	6.530	
%Stage 3	$\bar{X}$	7.125	6.650	.380*
	s	3.007	2.973	
%Stage 4	$\bar{X}$	12.510	13.295	.406*
	s	5.567	4.977	
%MT	$\bar{X}$	2.225	2.120	.451*
	s	1.527	1.192	
TST	$\bar{X}$	428.800	435.900	.127
	s	24.098	19.687	
WB1	$\bar{X}$	7.075	8.350	.198
	s	5.754	7.692	
WA1	$\bar{X}$	2.225	5.550	−.044
	s	2.967	10.050	
TB2	$\bar{X}$	14.050	10.393	.258
	s	9.434	18.875	
TBR	$\bar{X}$	81.025	92.875	.125
	s	23.852	25.522	
NuMT	$\bar{X}$	13.200	14.900	.620**
	s	7.951	7.691	
NuSC	$\bar{X}$	64.100	65.700	.558**
	s	24.859	24.862	

[a] Abbreviations Table 1 and Table 2: MT, movement time; TST, total sleep time; WB1, wake time before stage 1; WA1, wake time after stage 1; TB2, time to first stage 2 onset; TBR, time to first REM onset; NuMT, number of movements; NuSC, number of stage changes.

* $p < .05$, one-tail test.

** $p < .01$, one-tail test.

Results

The average total sleep time on night 3 was 428 min, with a range of 371 to 463 min; the average on night 4 was 435 min, with a range of 390 to 458 min. Averages of various sleep measures, as well as the average percentage of time spent in each sleep stage, are shown in Table 1. A zero-mu *t* test for correlated means on the 14 measures revealed no significant differences between night 3 and night 4. The Pearson

TABLE 2

Within-night correlations of sleep measures

Sleep Measures	Correlations	
	Night 3	Night 4
%REM × %Stage 2	−.304	−.126
%REM × %Stages 3 + 4	−.484*	−.015
%REM × %MT	.147	.144
%Stage 2 × %Stages 3 + 4	−.274	−.694**
%Stage 3 × %Stage 4	−.126	−.395*
TB2 × TMT	.066	−.110
%MT × %Stage 1	.344	−.046
NuMT × NuSC	.717**	.591**
%Wake × %MT	.113	.283
%Wake × %Stage 1	.130	.105

* $p < .05$, two-tail test.
** $p < .01$.

product-moment correlations for the 2 nights between *S*s are presented in the last column of Table 1. Significant positive correlation coefficients were found for percent sleep in stages 1, 2, 3, and 4, percent movement time, number of movements, and the number of stage changes. (Correlations were also calculated for time in min but gave identical results.) Rank-order rhos differed from the product-moment correlations by ±.14, indicating that deviations from bivariate normality were insignificant.

Within-night between-subjects Pearson product-moment correlations for 10 pairs of sleep measures are given in Table 2. The only correlation that was significant for both nights was that between the number of movements and the number of stage changes.

REM-NonREM Cycle Duration

The duration of the first REM cycle each night was measured in minutes from the onset of the first REM period to the onset of the second REM period. The duration of the second REM cycle of the night was measured from the onset of the second REM period to the onset of the third REM period. All 20 *S*s had at least 3 REM periods, but a few lacked a 4th and 5th REM period. The length of the REM period about doubled, from an average of 16 min for the first REM period to an average of 35 min for the fourth REM period. The duration of the first REM cycle was the shortest of the night, but there was no increase in REM cycle duration after the second REM cycle. In Table 3 are the durations of each *S*'s first and second REM cycles for nights 3 and 4.

There are six possible correlations among the four REM cycles. Only one was significant: the correlation between nights 3 and 4 for the first cycle is .42, significant at about the .03 level (one-tail test). The other correlations ranged from −.14 to .28, none reaching the .05 significance level (one-tail test). All correlations were recomputed using the Spearman rank-order method—but with the same result. Only the correlation for the first cycle was significant.

The usual measure of REM cycle duration is the average of all available cycle durations during a night; ordinarily, the average of several similar measures has a higher reliability than the uncombined scores. An average night 3 score was computed using the two to four cycle duration

TABLE 3

Duration of REM cycle in minutes for two cycles and two nights

Subject	REM Durations (min)			
	First REM Cycle		Second REM Cycle	
	Night B2	Night B3	Night B2	Night B3
02S68	80	87	76	76
03R68	134	120	162	90
04C68	70	91	116	94
05S68	109	108	90	118
06R68	82	71	94	120
07C68	77	64	92	86
08R68	80	94	158	90
09S68	84	104	60	67
10S68	83	112	111	82
11R68	58	56	96	100
12C68	80	86	60	112
13S68	113	92	107	120
01R70	70	105	97	84
02S70	107	83	76	88
03R70	74	81	107	156
04R70	74	62	99	84
05S70	76	122	110	100
06S70	94	72	124	102
07R70	86	98	106	72
08R70	80	62	116	89
Mean	85.6	88.5	102.8	96.5
Stand. Dev.	17.8	19.8	26.2	20.7

scores available for each *S*. An average night 4 score was calculated in the same way. The night 3 score had a mean of 92.1 min and a standard deviation of 17.3. The night 4 score had a mean of 92.8 and a standard deviation of 11.7. The correlation between nights 3 and 4 was .22, far below the .05 level of significance.

A within-subject within-night analysis of the REM cycle was done by comparing the duration of each REM period with the duration of the previous NREM period. Holding night and subject constant, this method yielded one correlation per subject. The mean of these correlations was −.206 for night 3 and −.402 for night 4. Both were significant at the .05 level by the signed-ranks test and the two-tailed zero-mu *t* test. An identical analysis was performed using the REM period paired with the following NREM period. The mean of the correlations was .046 for night 3 and −.176 for night 4. Neither was significantly different from zero by the *t* test or signed ranks test. Recall that during the night, the duration of the REM cycles tends to remain fairly constant. Since the REM period duration about doubles in the last third of the night, the NREM segment preceding the REM period must decrease; thus the negative correlation. But the length of that REM period cannot be used as an index of the length of the immediately following NREM period.

Discussion

Since Dement and Kleitman (1957) published their paper on cyclic variations during sleep and detailed a quantitative way of scoring sleep, these scores have been used as independent and dependent variables with surprising little concern over the reliability of these measures. The matter of constancy is usually handled as did Berger (1969) in an opening chapter in a symposium report on sleep: "The pattern of sleep from night to night in a single individual remains relatively constant, except for the first night spent in the laboratory when the subject takes a longer time to fall asleep, tends to awaken more frequently, and has less REM sleep than on subsequent nights [p. 21]." Except for the reports by Webb and Agnew (1969) and Johnson et al. (1970), there appear to be no published data to support this generally used and accepted statement. Our sample is comparable to the 16–19-year-old group studied by Webb and Agnew (1969). Only for stages 1, 2, and 4 were their retest correlations over 3 nights (.31, .60, and .48 respectively) statistically significant ($p <$ .05, one-tail). The correlations for wake and REM were not significant. Our correlation for stage 3 is higher than that found by Webb and Agnew, but the other five correlations agree. REM sleep, which has been the focus of much research, had much lower coefficients in both studies. Webb and Agnew reported lower coefficients for REM sleep than NREM sleep at all age levels, and the coefficient was not significant in the 30–39 age group.

Although most of our sleep stage retest correlations were significantly greater than zero, they, as well as those by Webb and Agnew, were far below the .7 to .8 needed for reliable predictions. Authors usually report the comparison of mean values over nights and note their similarity. Our mean scores also showed remarkable stability for every characteristic measured. Our correlations across subjects, however, indicate that the similar group values are very misleading with respect to night to night intrasubject consistency.

The consistent significant correlations reported in Table 2 for the number of movements and the number of stage changes show that frequently a stage change is accompanied by a movement and vice-versa. The number of movements and the number of stage changes have the highest correlations in Table 1. The relationship of body movements to stage changes has been previously noted by Sassin and Johnson (1968). They were also impressed with the consistency of the number of body movements over nights. The same consistency was noted by Johnson et al. (1970). Williams et al. (1964) reported stage change scores for 16 young adults over three nights. Their average intercorrelation is almost identical to our coefficient of .56. The average duration of the NREM-REM cycle was constant over nights, but the reliability of the REM cycle duration score was quite low. The set of six correlations calculated for the four duration scores does not differ significantly from that expected by chance. Our sample size, 20 *Ss*, was rather small, but if the true correlations are .5 or higher, our probability of detecting a positive correlation, significant at the .05 level, is .70 or better.

Could the reliability of REM cycles be improved by some other method than measuring the average interval from REM onset to REM onset? Globus (1970) uses an auto-agreement method which takes account of the REM-NREM status during every scoring period of the night. (He obtained an average REM cycle interval of 102 ± 9 min.) On the basis of his studies and others such as Weitzman et al. (1970), Globus proposed that relatively permanent aspects of the behavior of patients, such as psychotic behavior or depression, might be correlated with REM rhythmicity. We are currently trying to fit sinusoidal and boxcar (i.e., binary) periodic functions to binary REM scores as well as graded measures of EEG delta activity. These measures of REM cycle duration may well have greater reliability than the present method which relies heavily on estimating the exact time of REM onset. Until other methods are perfected, the low reliability of the current method of estimating REM cycle duration makes it a poor measure of any trait of the subject, though it may be a sensitive measure of the state of the subject. Our data further suggest that more attention should also be given to the reliability of other sleep measures before they are similarly used as stable individual traits.

REFERENCES

Agnew, H. W., Webb, W. B., & Williams, R. L. The first night effect: An EEG study of sleep. *Psychophysiology*, 1966, *2*, 263-266.

Berger, R. J. The sleep and dream cycle. In A. Kales (Ed.), *Sleep, physiology and pathology.* Philadelphia: J. B. Lippincott, 1969. Pp. 17-32.

Dement, W., & Kleitman, N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography & Clinical Neurophysiology*, 1957, *9*, 673-690.

Globus, G. G. Quantification of the REM sleep cycle as a rhythm. *Psychophysiology*, 1970, *7*, 248-253.

Hartmann, E. The 90-minute sleep-dream cycle. *Archives of General Psychiatry*, 1968, *18*, 280-286.

Johnson, L. C., Burdick, J. A., & Smith, J. Sleep during alcohol intake and withdrawal in the chronic alcoholic. *Archives of General Psychiatry*, 1970, *22*, 406-418.

Rechtschaffen, A., & Kales, A. (Eds.) *A manual of standardized terminology, techniques, and scoring system for sleep stages of human subjects.* Washington, D.C.: Public Health Service, U. S. Government Printing Office, 1968.

Sassin, J. F., & Johnson, L. C. Body motility during sleep and its relation to the K-complex. *Experimental Neurology*, 1968, *22*, 133-144.

Webb, W. B., & Agnew, H. W. Measurement and characteristics of nocturnal sleep. In L. E. Abt & B. F. Reiss (Eds.), *Progress in clinical psychology.* New York: Grune and Stratton, 1969. Pp. 2-27.

Weitzman, E., Kripke, D., Goldmacher, D., McGregor, P., & Nogeire, C. Acute reversal of the sleep-waking cycle in man. *Archives of Neurology*, 1970, *22*, 483-489.

Williams, R. L., Agnew, H. W, & Webb, W. B. Sleep patterns in young adults. *Electroencephalography & Clinical Neurophysiology*, 1964, *17*, 376-381.

Williams, R. L., Agnew, H. W., & Webb, W. B. Sleep patterns in the young adult female: An EEG study. *Electroencephalography & Clinical Neurophysiology*, 1966, *20*, 264-266.