Eur Respir J 1998; 11: 458–461
DOI: 10.1183/09031936.98.11020458

European Respiratory Journal
ISSN 0903 - 1936

Validation of automated sleep analysis in normal children

M.P. Villa, S. Piro, A. Dotta, E. Bonci, P. Scola, B. Paggi, M.G. Paglietti, F. Midulla, R. Ronchetti

Validation of automated sleep analysis in normal children. M.P. Villa, S. Piro, A. Dotta, E. Bonci, P. Scola, B. Paggi, M.G. Paglietti, F. Midulla, R. Ronchetti. ©ERS Journals Ltd 1998.

ABSTRACT: With the aim of determining normal reference values for our sleep laboratory and evaluating the reliability of automated analysis for scoring polysomnographic studies in children, we recorded polysomnograms in 16 healthy boarding-school children.

Sleep recordings were obtained with a computer system (Medilog SAC, Oxford Instruments). Polysomnographic variables were monitored continuously on a 16-channel recorder equipped with a video. Data were acquired on optical disk for computer-assisted data interpretation. Sleep stages and respiratory events were also scored visually by operator.

Comparison with visual scores showed that the computer system significantly overscored wakefulness (W) ($p<0.02$) and stage IV ($p<0.001$) and underscored stage II ($p<0.001$) and rapid eye movement (REM) sleep ($p<0.001$). It also assigned respiratory events a higher score than did visual scoring, as shown by the higher apnoea index (AI) and hypopnoea index (HI) (AI $p<0.03$; HI $p<0.001$). Regression analysis showed a significant correlation between visual and automated scores for central ($r=0.679$; $p<0.004$) and obstructive apnoea ($r=0.631$; $p<0.008$). Computer apnoea scores did not correlate with visual scores.

Much remains to be done before computer-based scoring systems can be relied upon, without visual scoring, for polysomnographic sleep studies in children. Their main advantage at present is that they offer a convenient means of saving paper, space and time.

Eur Respir J 1998; 11: 458–461.

II Cattedra di Clinica Pediatrica, Università "La Sapienza", Rome, Italy

Correspondence: R. Ronchetti
II Cattedra di Clinica Pediatrica
Università "La Sapienza"
Viale Regina Elena, 324
00161 Rome
Italy
Fax: 39 64462767

Keywords: Automated analysis
children
sleep

Received: March 6 1997
Accepted after revision August 25 1997

The study was presented in part at the Meeting of the American Thoracic Society 1994.

Polysomnography has a prime role in the study of children and adolescents with sleep-related breathing disorders [1]. Although the criteria for evaluating and analysing respiratory events during sleep have been standardized for adults and infants, few reported studies refer to the paediatric age [2–4]. Reference values are also scarce for identifying sleep stages in children from the age of 1 yr to puberty [5]. In a recent report, MARCUS *et al*. [3] have provided valuable data about normal values of respiratory events in children and adolescents. New computer-assisted scoring systems developed in the past decade have helped to make sleep studies simpler and less time consuming. But these methods have yet to be optimized even for use in adults. As ANDREAS *et al*. [6] confirmed in a previous paper, the results of automated analysis of sleep stages and apnoea need checking by visual scoring. A visual check is even more essential in sleep recordings of infants and children. The first reason is the high frequency of recording artefacts in children; the second is the current lack of paediatric reference samples for computerized polysomnographic systems. With the aim of determining normal reference values for our sleep laboratory and evaluating the reliability of automated analysis for scoring polysomnographic studies in children, we recorded polysomnograms in 16 healthy boarding-school children at the Institute "S. Anna" in Crotone.

Materials and methods

Study subjects

Sixteen normal healthy children, eight males and eight females, ranging 6–10 yrs in age (weight = 27.8±3.9 kg, height 131.9±6.4 cm) were studied. All were boarders (September to June), at the Institute S. Anna Crotone (Catanzaro, Italy).

None of the children had a history of sleep apnoea referred by parents or tutors by questionnaire, none were taking drugs; all subjects were healthy at the time of the study. One child found that the polysomnographic recording interfered with falling asleep; according to the questionnaire two children were occasional snorers.

Methods

Sleep recordings were obtained in the buildings where the children usually lived. As a sleep laboratory we used a quiet, comfortable, temperature-controlled room (22°C) equipped with a bed collected to a computer system (Medilog SAC, Oxford Instruments, Oxon, UK). To avoid disturbing the children's quality of sleep, the technician supervised the monitors from outside the room, through a

window. Testing usually began at 21:00 after a free-diet meal. No sedation or sleep deprivation were used. Overnight sleep studies were recorded with standard polysomnographic techniques [7].

Monitoring included electroencephalogram (EEG) (C3/A2; C4/A1); electro-oculogram (EOG1, EOG2); submental electromyogram (chin EMG); anterior tibial electromyogram (leg EMG); electrocardiogram (ECG); impedance plethysmography for recording chest and abdomen respiratory movements; and oronasal airflow *via* a thermistor. Oxygen saturation (S_{a,O_2}) was evaluated by pulse oximetry with a digital probe, recording S_{a,O_2} values and not the pulse wave. Polysomnographic variables were monitored continuously on a 16-channel recorder equipped with a monitor, and data were acquired on optical disk for computer-assisted data interpretation.

During the test, an operator checked for qualitative signal changes induced by the subject changing position or an electrode becoming disconnected.

Standard waveform detection criteria were used for computerized staging [8].

The computer-based system analysed polysomnographic recordings in real time and acquired the data on optical disk. On the following day, the operator reviewed the polysomnographic data and assigned a visual score. In four randomly selected subjects a second, blind operator reviewed the polysomnographic data to validate the first operator's scores. One scorer was a paediatric pulmonologist, the other a paediatrician expert in respiratory sleep disorders.

The computer-based system evaluated the following polysomnographic variables. Obstructive apnoeas were defined as the presence of chest wall and abdomen motion associated with the absence of airflow detection by the thermistor at the mouth and nares. The number and duration of obstructive apnoea lasting >5 s were quantified according to paediatric criteria [3–7, 9].

Central apnoeas were defined as absence of chest-wall and abdomen motion associated with absence of airflow detection by the thermistor at the mouth and nares. Central apnoeas lasting >5 s were included in the score, even if long central apnoeas are common in normal children [10] and do not appear to have clinical sequelae [11].

Mixed apnoeas were defined as apnoea having both central and obstructive components, when an absence of respiratory movements followed obstructive apnoea or *vice versa*, and the central component lasted >5 s. All respiratory events were quantified for whole recordings and subdivided into rapid eye movement (REM) and nonREM (NREM) stages.

The apnoea index(AI) was defined as the number of apnoeas per hour of sleep. Even if standard criteria for defining hypopnoeas in children have not been established, we considered hypopnoea as a >50% drop in the airflow signal associated with paradoxical inspiratory rib cage motion.

The mean and lowest S_{a,O_2} and the number of desaturations >4% were quantified. Measurements associated with poor pulse tracings were discarded. Total sleep time (TST) was quantified in minutes; sleep stages were expressed as a percentage of TST and scored in 30 s epochs according to standard criteria [7].

All children completed a questionnaire on psychophysical status, daily activities and possible use of drugs a few hours before the polysomnographic study. On awaken- ing the children also answered questions about what they thought of the procedure and whether the computer had interfered with their sleep.

Statistical analysis

The differences between human and computer-based scoring were calculated for all epochs (number of epochs and per cent of coincidence), and for each single epoch. Data are expressed as mean±SD and were processed by the statistical program SPSS/PC PLUS 6 using analysis of variance (ANOVA), two-tailed Wilcoxon matched-pairs test, Fisher's F-test and kappa statistic when appropriate. The relationship between two variables was assessed using linear regression analysis. A p-value of less than 0.05 was considered significant.

Results

Comparison between automated and visual scoring for sleep stages (expressed in per cent of the TST) showed that the computer system significantly underestimated TST ($p< 0.007$), overscored wakefulness (W) ($p< 0.02$) and stage IV ($p<0.001$) and underscored stage II ($p<0.001$) and REM ($p<0.001$) (table 1). No correlation was found between computer-assisted and visual scores for stages II, IV and REM but both methods yielded similar scores for stage III ($r=0.661$; $p<0.005$) (table 1). Calculation of the per cent agreement rate between computer and visual analysis using epoch-by-epoch comparison showed the highest agreement rate (78%) for stage IV and the lowest agreement for stages W, I and REM (table 2).

Table 1. – Computer-assisted analysis and visual sleep staging during polysomnographic recording in children

Stages	Sleep stage % of TST						TST min
	W	I	II	III	IV	REM	
Computer scoring	23.4 (19.2)	5.5 (3.2)	10.3 (4.4)	13.4 (9.5)	47.0 (9.5)	0.4 (0.7)	310.66 (108.34)
Visual scoring	9.5 (8.4)	4.8 (3.1)	31.1 (8.0)	18.4 (3.3)	26.7 (5.2)	16.5 (4.8)	396.44 (52.64)
p-value	<0.02	NS	<0.001	NS	<0.001	<0.001	<0.007

Values are expressed as mean, with standard deviation in parentheses. W: awake; REM: rapid eye movement. TST: total sleep time; NS: nonsignificant.

Table 2. – Epoch-by-epoch comparison between computer-based and visual scoring of sleep stages

Computer analysis	Visual analysis						
	W	I	II	III	IV	REM	Row total
W	1588	186	1291	249	625	555	4494
I	45	144	25	7	1	4500	672
II	70	145	284	85	39	636	1259
III	31	78	751	333	400	436	1669
IV	24	58	1669	1129	2601	298	5779
REM	3	0	0	0	0	40	43
Column total	1761	661	4020	1803	3306	2415	13916

Row total indicates the total number of epochs scored by computer analysis for each stage. Column total indicates the total number of epochs scored by visual analysis for each stage. For definitions see legend to table 1.

Table 3. – Respiratory events detected by computer scoring and visual scoring in a polysomnographic study

	Apnoea index	Hypopnoea index	Apnoea index NREM	Apnoea index REM	Hypopnoea index REM
Computer scoring	3.17 (2.00)	18.93 (11.55)	2.38 (2.11)	15.26 (34.51)	77.11 (180.04)
Visual scoring	2.06 (1.03)	3.29 (2.16)	1.31 (0.68)	2.86 (2.87)	6.00 (5.28)
p-value	<0.03	<0.001	<0.005	NS	NS

Values are expressed as mean, with standard deviation in parentheses. Data were analysed by analysis of variance. REM: rapid eye movement; NREM: nonREM; NS: nonsignificant.

The computer system assigned respiratory events a higher score than did visual scoring, as shown by the higher AI and hypopnoea index (HI) (AI p<0.03; HI p<0.001) (table 3). Apnoea detected and scored by the computer system did not correlate with apnoea scored visually. Linear regression analysis showed a significant correlation between visual and automated scoring for both central (r=0.679; p<0.004; slope=1.31; intercept=1.95) and obstructive apnoea (r=0.631, p<0.008; slope=1.07; intercept=1.42); and also for hypopnoeas (r=0.594; p=0.02; slope=3.17; intercept=8.49).

The first operator and blind operator scores agreed for all sleep stages and respiratory events (table 4).

The microphone recording obtained during polysomnography showed that the two occasional snorers identified by the questionnaire did not snore during the study.

Discussion

The primary aim of our study was to compare the performance of a computer-assisted system and visual analysis in scoring neurological and respiratory variables detected by polysomnography in children. We also sought normal reference values for our sleep laboratory. On this point our data, even if obtained in a small population, agreed with previously published data [3].

As early as 1971, in a comparison of computer-assisted and visual sleep scoring systems SMITH *et al.* [12] and MARTIN *et al.* [13] pointed out that computer scoring showed a disagreement, above all for stages III and REM. Subsequently, HOLLER and RIEMER [14], confirming the nonagreement between automatic and visual REM stage scores, concluded that technical advances in computerized systems would ultimately solve these problems. In a reevaluation of sleep scoring systems, KUBICKI *et al.* [15] reached the same conclusion, albeit computerized systems were becoming more reliable.

In a detailed study, ANDREAS *et al.* [6] have also addressed polysomnographic scoring of apnoeas and hypopnoeas, concluding that despite the 20 yrs that have passed since their introduction, computerized systems do not yet provide wholly reliable data. Computer-assisted systems markedly underscore stage I sleep and REM; and overscore stage II. Neither do they assess respiratory events reliably. They underestimate apnoeas, possibly because of the low sensitivity of thermistor in detecting respiratory flows. Finally, ANDREAS *et al.* [6] concluded that automa-ted analysis should be used only by those who are able and willing to perform a visual analysis.

The computer-based scoring system in our study achieved worse results than systems in other studies, because of the lack of agreement for both sleep staging and respiratory events. Our automated system failed to detect the REM stage because it confused the sleep-related movements typically seen in children with awakenings during sleep, thus overestimating awakening. It also failed to distribute the other sleep stages appropriately. It tended to identify stage IV better than the other stages. These uneven results probably depended chiefly on the lack of paediatric reference data for the computer system. Another important reason why the two scoring systems differed in performance is the low sensitivity of the methods used for detecting respiratory events (thermocoupled and impedance plethysmography). In evaluating children's respiratory events, the computer system emphasized their severity; the ratio between automated analysis and visual detection being 1.22:1.0 for apnoea and 5.2:1.0 for hypopnoea. The computerized system produced equally unreliable results for all respiratory events tested: central or obstructive apnoea, and hypopnoea. Because children produce lower signal variation in oronasal airflow and respiratory movements than adults, the low sensitivity of thermistor and impedance plethysmography certainly influenced the overestimation of respiratory events. If a subject's movement causes the thermistor to become dislodged, a computer-based system might falsely detect long apnoeas not associated with oxygen desaturations.

In summary, much remains to be done before computer-assisted systems can be relied upon, without visual scoring, for polysomnographic sleep scoring in children. Their main advantage at present is that they offer a convenient means of saving paper, space and time.

Table 4. – Agreement between two human scorers in detecting respiratory events and sleep stages

	Subjects n	TST min	Sleep stage % of TST						Apnoea index	Hypopnoea index
			W	I	II	III	IV	REM		
Operator	4	401.75 (73.24)	8.67 (10.84)	7.27 (4.10)	29.10 (6.45)	12.02 (1.34)	28.47 (3.34)	14.47 (6.67)	2.42 (0.98)	3.10 (2.06)
Blind operator	4	406.25 (76.78)	6.95 (11.90)	7.17 (0.56)	24.52 (2.50)	20.07 (6.66)	23.40 (5.72)	17.90 (9.01)	2.40 (0.89)	3.5 (3.9)
kappa		0.96	0.93	0.32	0.7	0.90	0.77	0.95	0.96	0.90
p-value*		<0.001	<0.003	<0.19	<0.09	<0.04	<0.09	<0.01	<0.02	<0.02

Values are expressed as mean, with standard deviation in parentheses. For definitions see legend to table 1.

References

1. Brouillette RT. Assessing cardiopulmonary function during sleep in infants and children. *In*: Beckerman RC, Brouillette RT, Hunt CE, eds. Respiratory Control Disorders in Infants and Children. Baltimore, Williams & Wilkins 1992; pp. 125–141.
2. Brouillette RT, Weese-Mayer DE, Hunt CE. Disorders of breathing during sleep in the pediatric population. *Semin Respir Med* 1988; 9: 594–606.
3. Marcus CL, Omlin KJ, Basinki DJ, *et al*. Normal polysomnographic values for children and adolescents. *Am Rev Respir Dis* 1992; 146: 1235–1239.
4. Guilleminault C, Van Den Hoed J, Mitler MM. Clinical overview of the sleep apnea syndromes. *In*: Guilleminault C, Dement WC, eds. Sleep Apnea Syndromes. New York, Liss, 1978; pp. 1–123.
5. Anders TF, Sadeh A, Appareddy V. Normal sleep in neonates and children. *In*: Ferber R, Kryger TM, eds. Principles and Practice of Sleep Medicine in the Child. Philadelphia, Saunders Co, 1995; pp. 7–18.
6. Andreas S, Von Breska B, Magnusson K, Kreuzer H. Validation of automated sleep stage and apnea analysis in suspected obstructive sleep apnea. *Eur Respir J* 1993; 6: 48–52.
7. Rechtschaffen A, Kales A, eds. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles, National Institutes of Health Publication, l978.
8. Operator's Manual. Medilos Sac. Oxford, UK, Oxford Instruments Plc. 1991; pp. 113–114.
9. Marcus SL, Carrol JL, Bamford O, Pyzik P, Loughlin GM. Supplemental oxygen during sleep in children with sleep-disordered breathing. *Am J Respir Crit Care Med* 1995; 152: 1297–1301.
10. Poets CF, Stebbens VA, Samuels MP, Southall DP. Oxygen saturation and breathing patterns in children. *Pediatrics* 1993; 92: 686–690.
11. Weese-Mayer DE, Morrow AS, Conway LP, Brouillette RT, Silvestri JM. Assessing clinical significance of apnoea exceeding fifteen seconds with event recording. *J Pediatr* 1990; 117: 568–574.
12. Smith JR, Karacan I. EEG sleep stage scoring by an automatic hybrid system. *Electroenceph Clin Neurophysiol* 1971; 31: 231–237.
13. Martin WB, Johnson LC, Viglione SS, Naitoh P, Joseph RD, Moses JD. Pattern recognition of EEG-EOG as a technique for all-night sleep stage scoring. *Electroenceph Clin Neurophysiol* 1972; 32: 417–427.
14. Holler L, Riemer H. Comparison of visual analysis and automatic sleep stage scoring (OXFORD Medilog 9000 System). *Eur Neurol* 1986; 25 (Suppl. 2): 36–45.
15. Kubicki ST, Holler L, Berg I, Pastelak-Price C, Dorow R. Sleep EEG evaluation: a comparison of results obtained by visual scoring and automatic analysis with the Oxford sleep stager. *Sleep* 1989; 12 (2): 140–149.