Investigations of an automatic screening device (MESAM) for obstructive sleep apnoea

R. Stooohs, C. Guilleminault

Investigations of an automatic screening device (MESAM) for obstructive sleep apnoea. R. Stooohs, C. Guilleminault.

ABSTRACT: A digital recording device developed to monitor heart rate (HR) and breathing sounds (snoring), and used to screen subjects for obstructive sleep apnoea syndrome (OSAS), was investigated. This device is called the MESAM and is commercially available in some western European countries. The computer-based automatic scoring systems provided with the equipment and a hand-scoring technique developed at Stanford and requiring 10-15 min to perform, were used. Polysomnography and MESAM recordings were performed simultaneously on two groups of 25 sleep disorder patients (each with respiratory disturbance Index > or < 10). Patients were randomly monitored and records were analysed by two teams blind to the initial clinical impression, to the events which occurred during the recordings, and to each other’s findings. Specificity and sensitivity were calculated for each of the MESAM scoring techniques, considered, with polysomnography being selected as the recording standard. With "automatic HR scoring" specificity was 12%, sensitivity 92%; with "automatic breathing sounds (snoring)" scoring, specificity was 8% and sensitivity 96%; with "hand-scoring" specificity was 72% and sensitivity 92%. If the three scoring techniques were combined, all patients with a respiratory disturbance index (RD) >10 were recognized as having OSAS.


A recent international symposium on obstructive sleep apnoea syndrome (OSAS) revealed strong philosophical differences concerning the identification of OSAS patients. North American and Australian researchers favoured labour-intensive polygraphic monitoring, whilst European clinicians favoured the development and usage of screening devices [1]. These devices could be used in private practice to perform a qualification process which would select patients for referral to specialists for further investigation. Several devices are currently commercially available in various countries in western Europe [1]. Their function is to assure the physician of the absence of a clinically relevant obstructive sleep apnoea problem in a patient who presents a history of sleep disorder and/or snoring.

The intention in using these devices is to reduce screening costs and avoid inappropriate nocturnal polygraphic recordings. We have performed a validation study of one of these commercially available devices: the MESAM. We compared the results obtained with this screening device with those tabulated from standard polysomnography.

The equipment

MESAM is a digital recording device developed by Madaus Medizin Elektronik and the University of Marburg Sleep Laboratory (FRG) to monitor heart rate (HR) and breathing sounds. Since 1987, there have been several publications describing the device [2-5]. To summarize: a small box, 30 mm x 90 mm x 160 mm, with a total weight of 365 g, including four AA 1.5 v batteries, contains electronic circuit boards. Heart rate (HR) is monitored through a single-lead ECG (modified D), and R-R intervals are measured in ms. The time interval is instantaneously compared to the value placed in the permanent memory of the devices. Snoring sounds are monitored through an electric, subminiature microphone, type MCE 2,000 (frequency range 30-20,000 cycles·s⁻¹ ±2dB, sensitivity 0.6 mV·µ bar⁻¹ at 1,000 cycles·s⁻¹ ±4dB, available from Conrad electronics, Hirschau, FRG). The microphone is encapsulated in such a way that it stands 2 mm from the skin to avoid the noise caused by rubbing, and it is taped above the larynx [3-5]. The 2 mm small "pit" communicates through a
0.5 mm channel that allows air movement, so as to avoid pressure changes related to neck movement directly on the membrane of the microphone.

Snoring sounds have been defined, based on the power spectrum analysis of sounds [3–5]. If the relative power of the frequencies between 50–800 cycles exceeds 50% of the total power, "snoring" is identified and one bit is set in the MESAM device. Another bit takes total power of the sound into account. If the volume exceeds a predefined threshold corresponding to "loud snoring", which is defined by high microphone output (>1.1 mV at 1,000 cycles·s⁻¹), another bit is set. Eight bits of information are collected every second, and they are stored in a solid state memory (64 kilobytes of RAM). The beginning of the recording is programmable; thus attachment of the device is independent of the start of the recording. An IBM-compatible personal computer is used for data analysis. A graphic card provided with the equipment is used for the automatic analysis of HR and breathing sounds and the printing of data on hard copy. The algorithms used for automatic analysis are partially described in publications of the Marburg University research group [2–5] but are not given by the company at the purchase of the equipment. The Marburg group publications have emphasized that MESAM can be hand-scored by 10 min intervals using the degree of variation of HR with the help of full resolution tracings, called "overview". However, strict criteria for hand-scoring were unavailable at the time of our study and have not yet been published.

Patients

A total of 50 patients were included in the study. The protocol called for investigation of 25 patients with symptoms of obstructive sleep apnoea syndrome (OSAS) and 25 patients with suspicion of other sleep disorders at clinical interview. Clinical interviews were performed by physicians not involved in the study. They classified patients, based on their clinical impressions, following the nosology from the Association of Sleep Disorders Centres (ASDC) [6]. The subject to be equipped with MESAM on a given night was selected at random within these two groups by a physician not involved in the study. The night of the recording was determined by the availability of the equipment. All data analyses were performed blindly.

Methods

Patients were asked to be in the sleep clinic by 19.00 h. They were provided with recording equipment and MESAM between 19.00 and 21.00 h. According to their preference and personal routine, patients retired between 21.30 and 23.30 h and were awakened at 06.00 h the next morning.

Polysomnography

The following variables were systematically monitored: EEG (C3/03; C4/04), electro-oculogram, chin and leg electromyograms and ECG (modified D3 lead). Respiration was monitored by uncalibrated Respiratime (thoracic and abdominal effort). Airflow was monitored by thermistors. Arterial oxygen saturation (Sao2) was monitored by Biox-Ohmeda or Nelcor oximeters. MESAM was installed simultaneously with other sensors. Lights-out and lights-on times were indicated on the MESAM information sheet. Technicians performed polygraphic recordings and obtained MESAM data, which were then scored in a blind fashion by two different teams whose members had not been involved in the polygraphic recording. A polysomnographic scoring team consisted of two trained technicians who scored sleep/wake, apnoeas, hypopnoeas and Sao2 drops, following international criteria for scoring sleep and respiration. Sleep/wakefulness was scored following the criteria of RECHTSCHAFFEN and KALES [7], and apnoeas and hypopnoeas were scored according to the criteria of GUILEMINAULT et al. [8], being subdivided into central, obstructive, and mixed events. An apnoea is defined as an interruption of air exchange lasting 10 s or longer, and a hypopnoea as a drop in airflow recording of at least 50%, associated with a change in Sao2 from baseline. Sleep, respiration and periodic leg movements were analysed by 30 s epochs. These are the criteria currently accepted by the international community.

MESAM

MESAM was analysed using the two automatic machine-based scoring systems, i.e. snoring-based and HR-based. It was also scored, using strict pre-established criteria, by one investigator, blind to patient status and to polysomnographic findings. The hand-scoring was performed in 10 min segments, as recommended by previous publications [3–5], which allowed temporal identification of scored events and appropriate comparison with temporal scoring of the polysomnogram. The hand-scoring of MESAM followed the pre-established criteria outlined below. In the absence of a published manual for hand-scoring or clear criteria from scientific publications, we performed an open trial of the equipment with simultaneous polygraphic recording on 10 subjects. We established the following definitions for use during the study.

Pre-established criteria for hand-scoring of MESAM. The hand-scoring was performed on the so-called "full disclosure" recording, with HR recording divided into 10 min epochs (fig. 1).

Regular HR at the beginning of the night was determined. Using the patient sleep log, lights-out time and lights-on time were determined. If the log contained information from the patient concerning period(s) of awakening during the night and timing of lights-out and lights-on, these periods were logged on the full
Fig. 1. – Recording from MESAM, example of a "full disclosure" record on which hand-scoring was performed. Each horizontal segment represents 10 min of recording. Time is indicated twice for each segment. Breathing sounds (snoring) are on the top and heart rate on the bottom of each segment. Dark blocks on the "snoring" line represent snoring period; if no snoring occurred, as can be seen during most of the first 9 min (top of the figure), the print-out is blank. R-R intervals are presented in bpm (scale on the left-hand side of figure) for each 10 min segment.
Disclosure MESAM print-out. Heart rate just before lights-out and after lights-on was considered as “wake HR”.

Each 10 min epoch was analysed independently. During the 10 min epoch, each snore was identified and snoring intervals (i.e. time between each snore) were noted. A pattern in which snores are separated by silent intervals of 10–120 s duration suggests an apnoea. When an apnoea was suspected an analysis of HR was then performed to confirm the classification of apnoea/hypopnoea. This analysis was performed on the HR at its change during the snore and during the silent period just preceding and just following the snore. During the silent breathing in between each snore a baseline HR was defined: it was the regular HR preceding and/or following any abrupt increase or decrease in HR.

Changes in HR from baseline were then analysed. Each HR peak above 10 b·min⁻¹ from the baseline HR was counted as “apnoea/hypopnoea” [9]. However, the HR change must present a “peak”, not a “plateau”, at maximum increase of HR, to be considered related to an "abnormal respiratory event". An HR >35 b·min⁻¹ over HR baseline was considered to be an arousal. If there was a change in HR ending with a clear peak (i.e. with increase and decrease), but without reaching a 10 b·min⁻¹ increase, this peak was also considered indicative of an abnormal respiratory event, not only if it occurred during a silence between two snores but also if the peak HR fell in the interval between snores. Snoring was scored as “loud” or “soft”, based upon the width of the individual snore print-out [1]. This width was defined by the MESAM software.

If the baseline HR changed abruptly, increased by at least 5 b·min⁻¹ and maintained the new rate, this was considered indicative of an awakening. If the increase presented a limited plateau and returned to a lower level, this was considered an arousal. Sleep was considered to occur when HR dropped from baseline (just prior to lights-out) level.

Results

MESAM is a safe, easy-to-use, simple device. Once a microphone had to be replaced because it had been broken by a patient pulling on the wire to remove the equipment. No other dysfunction was noted.

Polygraphic recording (PSG)

Some patients with other sleep disorders presented few sleep apnoea events. However, patients suspected of having OSAS always presented a respiratory disturbance index (RDI) (apnoea + hypopnoea index) >10. We can thus define a group of 25 patients with clinical symptoms of OSAS and an RDI>10 (Group A) (mean age 49.8±11.2 yrs, mean body mass index (BMI) 28.1±5.0 kg·m⁻² and a group of 25 subjects predominantly with other sleep complaints and an RDI <10 (Group B) mean age 29.2±4.3 yrs (Ns), mean BMI 29.2±4.3 kg·m⁻² (Ns).

Classification of subjects in Group A or B

False negatives in OSAS Group A. No subject was misclassified using the three scoring techniques for MESAM. However, using two scoring criteria two subjects were not properly identified: using PC scoring software, these two subjects were false negatives with HR algorithm, but only one subject was a false negative with snoring algorithm (4%). One subject was mis-classified with hand-scoring criteria (4%), but was properly classified by the snoring algorithm. The sole usage of the automatic scoring would thus have missed one subject entirely. Usage of hand-scoring alone would have missed a different subject. This subject had been identified as a snorer without apnoea by hand-scoring, and the PC software had in fact classified him/her as an apneic subject because of the snoring pattern.

False positives in Group B. Seven subjects out of 25 were misclassified even when combining the three methods. Automatic scoring misclassified 22 subjects out of 25 using HR, and 22 out of 25 using snoring. Seven subjects out of 25 (also not identified by automatic analysis) were misclassified by hand-scoring.

In summary MESAM had difficulty classifying subjects with few apnoeas. The algorithms selected for automatic analyses did a poor job. Hand-scoring criteria were better, but still misclassified seven subjects (fig. 2).

Investigation of misclassification. Further analysis was performed to investigate scoring discrepancies.

1. False negatives. Investigation of the two subjects involved in misclassification revealed that one patient had autonomic nervous system dysfunction; the other had predominantly hypopnoeas with only few apnoeas (but significant desaturation).

2. False positives. In Group B, 14 subjects (56%) presented periodic leg movement (PLM). Mean RD1so for these patients was 4. Automatic HR-scoring revealed a mean RDI of 35; automatic scoring of snoring intervals a mean RDI of 20. Mean hand-scoring results for this group was 10.

PLM is thus a difficult differential diagnosis with OSAS when using MESAM; more so if the subject snores without apnoea/hypopnoea.

We also investigated PLM in the OSAS group. There were nine cases with PLM (36%). Mean RD1so was 37; mean RDI of HR-score (RD1s) 52; mean snoring-RDI score 38; and mean hand-score RDI 34. When Groups A + B were analysed together subjects 23 and 50 (46%) presented PLM. Mean RD1so in this subgroup was 18; mean RD1s score 43; mean RDI of snore-score (RD1sn) 28; mean RDI of hand-score (RD1sh) was 21.

In summary, analysis of false positive results shows that PLM plays an important role in deviant results. PLMs can have an effect similar to an arousal induced by termination of apnoea, when the variables monitored with MESAM are analysed to determine presence or absence of breathing disorders during sleep.
Mean deviations from polygraphically determined RDI (RDI_{PSG})

We also analysed discrepancies between polygraphic recordings and MESAM scores. Firstly, we calculated the "delta" for each of the three scoring methods. This delta was calculated without regard for false negatives or false positives; we considered only the difference of RDI in absolute value. We then calculated the mean deviations from RDI_{PSG}, which were as follows: RDI_{10}=27.0; RDI_{30}=20.4; RDI_{90}=13.42.

Evaluation of the role of sleep-efficiency

Analysis of the raw data indicated substantial discrepancies in evaluation of total sleep time between MESAM and polygraphic recordings. To better evaluate the role of total sleep time determination in the observed differences, we performed the following analysis. As the RDI_{PSG} gave the best MESAM RDI/RDI_{PSG} relationship, it was selected as MESAM RDI. We looked at the patients who presented the largest delta (as defined above) compared to RDI_{PSG}.

A regression line was calculated for the delta between the two RDIs on the sleep-efficiency and is presented in figure 3. As can be seen, there is a very clear relationship between poor sleep-efficiency (i.e. closer to 0.00) and high delta value points. The r of the regression line is 0.95. This indicates that as sleep disruption and amount of time spent awake at night (and, thus, the difficulty of appreciating total sleep time appropriately with a simple device such as MESAM) increases, the discrepancy between RDI_{PSG} and RDI_{90} increases. With MESAM, total sleep time is estimated from lights-out at night, lights-on in the morning, and the time periods indicated as "wake" by the patient (urination, etc.). RDI is calculated from this estimated total sleep time, which may be quite different from the EEG determined total sleep time. The difficulty in appropriately appreciating time awake during the nocturnal period is probably responsible for the most important discrepancies. With polygraphic recording a better estimation of total wakefulness during sleep is made.

Specificity and sensitivity of MESAM

Correlation analyses were performed using correlation coefficient and Spearman rank correlation. Specificity and sensitivity of MESAM were calculated, taking PSG as the standard, and considering the different RDI scores obtained with both automatic scoring systems and criteria defined for hand-scoring [10].

For Group A (RDI >10) the correlation coefficient (a) and Spearman rank correlation (b) were always best when the MESAM record was hand-scored, an analysis which took 10-15 min, compared to the available automatic scoring systems. (Hand-scored RDI correlations: a) 0.68 b) 0.69; automatic HR-scoring: a) 0.59 b) 0.63; automatic snoring based scoring: a) NS b) NS). For Group B (RDI <10) there was no significant correlation.

If we consider the capability of MESAM to classify subjects into two groups (RDI >10 and RDI <10), we can calculate the specificity and sensitivity of the system for each of the MESAM scoring systems. Using the computer based HR-scoring capability the specificity was 12% and sensitivity 92%; using the computer based snoring scoring capability, the specificity was 8%; the sensitivity 96%; using the criteria developed to hand-score the MESAM record, the specificity was 72%, the sensitivity 92%. If the PC scoring software was considered alone, the sensitivity was 96% and the specificity was 20%. If our own pre-established criteria were used in association with the machine criteria, the sensitivity was 100% and the specificity 80%.
Regression of delta between RDI hand and RDI PSG on sleep-efficiency

Fig. 3. - Regression of the delta obtained when comparing hand-scored MESAM RDI to RDI$_{PSG}$ on sleep-efficiency calculated from PSG. As can be seen, when the MISSAM score is far apart from the RDI$_{PSG}$ there is a very good relationship between poor sleep-efficiency and high delta (r=0.95; R$^2$ adjusted=0.94). This indicates that the more important the sleep disruption and time awake during the night, the more the risk of a high delta. Difficulty in evaluating total sleep time with MISSAM is responsible for this discrepancy. For definitions see legend to figure 2.

Discussion

MESAM is commercially available, and has been used extensively by a well-known sleep centre at Marburg University, where it was initially developed. Undoubtedly the goal of a manufacturer is to sell his equipment to a large number of users. This allows him to develop a piece of equipment that is within a reasonable price range. Two other points in MESAM’s favour: 1) considering health policy, it is probably more economical to avoid inappropriate referrals to specialists; 2) in the recent past, many professionals have emphasized that sleep laboratories and clinics should not be flooded by “snorers”.

This is an easy-to-use piece of equipment and, because of its sturdiness and easy handling, it could undoubtedly be used by general practitioners with little knowledge of sleep and breathing disorders. The questions that we need to ask are these: 1) Is this equipment able to prevent inappropriate referrals to sleep centres without risk for the screened subjects? 2) Can it be useful in cutting costs, allowing general practitioners to feel that no further investigations are needed, without intervention of a sleep specialist? 3) Is this equipment more appropriate for the qualification of patients sent to a sleep specialist than for a general practitioner in his office? 4) Could this equipment be used for scientific studies (appreciation of obstructive sleep apnoea incidence in a given population [4], etc.)?

General practitioners will not have the sophistication of specialists. They may not obtain sufficient clues from a snorer to make an appropriate judgement concerning the need to refer him/her to a sleep centre. In this kind of situation the hand-scoring method will probably not be much used since it requires time and training. The PC software will probably be the basis of the final decision, as it can be handled perfectly well by a nurse or a technician. Considering the sensitivity results (96% at best), many but not all OSAS patients will be identified and, considering the specificity results, the equipment will do little to diminish the number of inappropriate OSAS referrals. As it stands, MESAM is probably not the complete answer for screening OSAS in private practice. When sensitivity has reached 100%, the automatic screening devices will probably be much sought-after. MESAM is currently being modified with the adjunction of continuous Sao$_2$ monitoring. This adjunction may bring automatic scoring to a sensitivity of 100%, but further equipment testing will be needed.

Would MESAM be helpful in a sleep centre? Usage of hand-scoring criteria would bring the sensitivity to 100% when subjects with RDI >10 were looked for (assuming that the centre wanted to isolate patients with RDI >10 and perform further evaluations on these patients). If this approach were used, considering the specificity, MESAM undoubtedly would identify more subjects than would polygraphic recording. This would not be a problem in a clinical setting, as patients would be recognized by the sleep laboratory evaluation. The specificity findings, in our view, would not make MESAM accurate in epidemiological investigations [4]; there would always be a clear discrepancy, with a risk of a one-sided bias, considering the number of false positives. However, as very few publications have come out regarding validation of questionnaires, non-PCGs, or other techniques used to perform epidemiological investigations, MESAM may be better than many other means. Lack of knowledge about the sensitivity and specificity of many other techniques undoubtedly gives MESAM an edge; at least now we have an indication of MESAM’s limitations. If researchers want to have a 3 or 4 point scale (i.e. absent, mild, moderate, severe), the results obtained here indicate that significant errors can be made. Furthermore, MESAM can only evaluate RDI; it cannot evaluate any of the multiple indices that have flourished over the years (arhythmia index, Sao$_2$ indices, etc.).

This investigation has clearly shown the difficulty in appreciating the number and length of short periods of wakefulness with MESAM. These difficulties will probably exist with most equipment which does not monitor sleep EEG. Actigraphy may help reduce the gap but, as shown in our study, if sleep efficiency is poor, RDI may be difficult to calculate. The margin of error will be important if apnoeas are intermittent and sleep efficiency poor for any reason. However, this type of patient is the kind that a sleep clinic or general practitioner may want to examine with MESAM-type equipment for presence or absence of apnoeas, before engaging in the treatment of important nocturnal sleep disruption, as there is as much reluctance to consider monitoring all insomnias as there is to monitor all reported snorers, whenever health economics are invoked.

In summary, MESAM may be helpful to physicians educated in sleep medicine, but its usefulness in general practice is limited at this point because of the possibility of false negative response with automatic scoring.
and because of the necessity of time-consuming hand-scoring analyses. The manufacturer is aware of this limitation in the equipment and the limits of the potential market; the adjunction of continuous Sao2 monitoring currently being done may make this portable equipment more desirable, but new validations of the modified system will be needed.

Acknowledgements: The authors thank Madaus Medizin Elektronik for providing the equipment and for their sense of fair play in the investigation of the system. They also thank L. Wilkinson for her help in editing the manuscript.

References


6. Association of Sleep Disorders Centers, Roffwarg HP (Chairman). Diagnostic Classification of Sleep and Arousal Disorders, 1st edn. Sleep, 1979, 2, 1–137.


RÉSUMÉ: Un appareil enregistrant le rythme cardiaque, et la présence ou l’absence d’un ronflement, et mettant en mémoire les données recueillies sous forme digitale (le MESAM), a été évalué. Cet appareil est utilisé pour la détection des sujets présentant un syndrome "d’apnée obstructive du sommeil". Les systèmes d’analyse automatiques vendus avec l’appareil et une analyse manuelle, qui demande 10 à 15 minutes de travail, basée sur des critères développés à Stanford, ont été utilisés. Les enregistrements polygraphiques nocturnes et les enregistrements avec MESAM furent réalisés simultanément sur deux groupes de 25 malades avec des troubles du sommeil et avec un index d’apnées et d’hypopnées inférieur ou supérieur à 10. Les enregistrements ont été fait au hasard, et analysés en aveugle, par deux équipes. La spécificité et la sensibilité de chaque technique d’analyse furent calculées par rapport aux résultats obtenus avec une polygraphie nocturne. Avec le système basé sur l’analyse automatique du rythme cardiaque, la spécificité était de 12% et la sensibilité de 92%. Avec l’analyse basée sur la respiration (ronflement) la spécificité était de 8% et la sensibilité de 96%; avec l’analyse manuelle, la spécificité était de 72% et la sensibilité de 92%. Si les 3 techniques étaient utilisées conjointement, tous les malades avec un index d’apnée-hypopnée >10 furent identifiés par MESAM.