Unreliability of Automatic Scoring of MESAM 4 in Assessing Patients With Complicated Obstructive Sleep Apnea Syndrome*

Fabio Cirignotta, MD; Susanna Mondini, MD; Roberto Gerardi, MD; Barbara Mostacci, MD; and Elisa Sancisi, MD

**Background:** Portable devices are used for unattended recording of patients with suspected obstructive sleep apnea syndrome (OSAS). The MESAM 4 (MAP; Martinsried, Germany) is a computerized ambulatory polysomnographic system that records four parameters: breathing noise, heart rate, arterial oxygen saturation ($\text{SaO}_2$), and body position.

**Design and method:** We evaluated the reliability of the oxygen desaturation index (ODI) automatically calculated by the MESAM 4 device in evaluating patients with “complicated” OSAS. These patients present $\text{SaO}_2$ drops due to apneas associated with a fall in baseline $\text{SaO}_2$ during sleep, as occurs in the “overlap syndrome.” Ten patients with complicated OSAS underwent nocturnal MESAM 4 recordings, and we compared the visual and automatic scorings of the ODI.

**Results:** The ODI obtained with visual scoring was significantly higher than ODI automatically calculated by the MESAM 4 in all patients. In some patients, this difference was so significant that it could bias clinical judgment of OSAS severity. We demonstrated that the system did not identify those desaturation events that were superimposed on a fall in baseline $\text{SaO}_2$. The error depends on the algorithm by which the device recognizes the desaturation events and calculates the baseline $\text{SaO}_2$.

**Conclusion:** Automatic analysis of MESAM 4 recordings may be misleading in evaluating OSAS patients who have a fall in baseline $\text{SaO}_2$ during sleep. In this case, visual scoring performed by a trained polysomnographer is recommended.

*CHEST 2001; 119:1387-1392*

**Key words:** cardiorespiratory monitoring; MESAM 4; sleep apnea

**Abbreviations:** ASDA = American Sleep Disorders Association; ODI = oxygen desaturation index; OSAS = obstructive sleep apnea syndrome; RDI = respiratory disorder index; $\text{SaO}_2$ = arterial oxygen saturation

In 1994, the American Sleep Disorders Association (ASDA) identified four levels of recordings for the assessment of obstructive sleep apnea syndrome (OSAS).¹ Level I or standard polysomnography is the attended in-laboratory recording of standard sleep and respiratory variables. Level II or comprehensive portable polysomography is the unattended off-site recording of the same variables as standard polysomnography. Level III or modified portable sleep apnea testing is the unattended off-site monitoring of a minimum of four cardiorespiratory variables. Level IV or continuous single-bioparameter or dual-bioparameter recording is the unattended monitoring of one or two cardiorespiratory variables. Levels III and IV do not allow for sleep scoring.

According the ASDA practice parameters, level II and III portable recordings are considered alternatives to standard laboratory polysomnography for the diagnosis of OSAS only in the following situations: (1) for patients with severe clinical symptoms indicating OSAS, when standard laboratory polysomnography is not readily available; (2) for patients who cannot be studied in the sleep laboratory; and (3) for follow-up studies to evaluate response to therapy.² Despite these restrictive guidelines, ambulatory systems have been widely used because they are less...
We analyzed MESAM 4 recordings of 10 patients with complicated OSAS (4 women and 6 men; mean age, 45.8 years). We classified patients as having complicated OSAS if they presented with sleep periods > 3 min in which rapid drops in SaO2 related to apneas were superimposed on a progressive and sustained fall in baseline SaO2 (Fig 1). The fall in baseline SaO2 could appear immediately after falling asleep, lasting part of the sleep time, or could be present for short periods during the night, possibly related to rapid eye movement sleep. All patients underwent a nocturnal recording with MESAM 4 for diagnosis of OSAS. The instrument was applied by a sleep technician in the laboratory 4 h before programmed starting of night recording at home. All standard MESAM 4 parameters (breathing sounds, heart rate, SaO2, body position) were monitored. In particular, oxygen saturation was monitored by a pulse oximeter with a sampling rate of 0.5 s. The sleep period to be analyzed for calculation of MESAM indexes was determined by the patient’s sleep log and indicated in the computer-based scoring system of the device.

Each MESAM 4 recording was scored blindly by two readers trained in polysomnographic interpretation. One scorer performed visual analysis of the recording on the printout of the “full-disclosure” mode in 10-min epochs. He analyzed the same sleep period automatically scored by the machine on the basis of sleep log. The desaturation events were visually identified by the following criteria: fall in SaO2 ≥ 4% compared with the previous value and a subsequent rise in SaO2 of at least 4%, with maximum duration of the event of 2 min. The visual ODI was calculated dividing the total number of events identified by the scorer by the hours of the analyzed period. Since the events automatically scored are underlined, visual analysis was performed before automatic analysis.

Another reader edited the automatic analysis. The ODI automatically calculated by the MESAM 4 device was obtained by dividing the total number of desaturation events recorded by the hours of the analyzed period. The system scores a “desaturation event” when SaO2 falls ≥ 4% from baseline; the event lasts until 95% of baseline has been reached again. The maximum possible duration of the event is 2 min, while a desaturation > 2 min is not scored as an event. Baseline saturation is calculated with an algorithm on the basis of the two highest values in the last 40 s.

The two scorers were instructed to read recordings blindly as part of a study on reliability of MESAM 4 automatic calculation of the ODI in OSAS patients, independently from their severity. The differences between visual and automatic scoring of the ODI were analyzed by paired t test and by the Bland-Altman procedure.

Results

In our group of patients with complicated OSAS who were studied with the MESAM 4 device, the ODIs automatically calculated were significantly underestimated with respect to the visual ODIs (mean ODI, 43.5 vs 75.1; p < 0.003). This difference was found in all patients; in some it was slight, while in others it was pronounced and could bias clinical judgment on OSAS severity. In particular, automatic ODI was in the range of mild-to-moderate OSAS in patients 1, 2, and 8, while visual scoring demonstrated very severe syndromes. Patient 7 also showed a marked difference (the visual ODI was twice the automatic score) even in the same range of severity. The difference between automatically scored ODI and visually scored ODI increased with the longer the sleep time with fall in SaO2 baseline. These
periods represented 34% of total sleep time as a mean, ranging between 5% and 79% (Table 1). Bland-Altman comparison of data indicated wide differences between methods of analysis (Fig 2). The fall in \( \text{SaO}_2 \) baseline occurred independently of body position.

The difference between manual and automatic analysis is due to the fact that the system does not score the desaturation events superimposed on a fall in baseline \( \text{SaO}_2 \), when this fall appears after the start of the automatic analysis (Fig 3). This failure depends on the algorithm by which the device recognizes a desaturation event and calculates the baseline. In fact, the software identifies the beginning of an event as a drop in \( \text{SaO}_2 \) of \( \geq 4\% \), and the end of an event as the point when \( \text{SaO}_2 \) returns to 95% of baseline; the return to baseline should occur within 2 min from the start; both components are necessary for an event to be scored automatically. In complicated OSAS, 95% of baseline \( \text{SaO}_2 \) is not always reached again with resumption of breathing, so the event cannot be terminated and is not identified by the MESAM 4 device.

The second error in the MESAM 4 algorithm is due to a software defect in calculating baseline \( \text{SaO}_2 \). The machine calculates the baseline at the time set by the scorer as “start time” of the automatic analysis (Fig 3). This failure depends on the algorithm by which the device recognizes a desaturation event and calculates the baseline. In fact, the software identifies the beginning of an event as a drop in \( \text{SaO}_2 \) of \( \geq 4\% \), and the end of an event as the point when \( \text{SaO}_2 \) returns to 95% of baseline; the return to baseline should occur within 2 min from the start; both components are necessary for an event to be scored automatically. In complicated OSAS, 95% of baseline \( \text{SaO}_2 \) is not always reached again with resumption of breathing, so the event cannot be terminated and is not identified by the MESAM 4 device.

The second error in the MESAM 4 algorithm is due to a software defect in calculating baseline \( \text{SaO}_2 \). The machine calculates the baseline at the time set by the scorer as “start time” of the automatic analysis. After this point if the \( \text{SaO}_2 \) drops by at least 4% and then fails to return to 95% of baseline values (i.e., if the program identifies the start of the event, but not its end), the system does not calculate a new baseline value, missing the scoring of superimposed desaturation events. This occurs independently of the rate
and degree of decline of the baseline $\text{SaO}_2$. There appears to be no time limit beyond which the machine recalculates a new baseline if the $\text{SaO}_2$ value fails to return to the original baseline. If we set the start time of the automatic analysis at a different recording point, when baseline was already lowered, a new baseline was calculated and, as long as it remains stable, the events previously unrecognized were automatically identified (Fig 4). However, we found a good correlation between visual and automatic scoring of the ODI when we considered series of desaturation events with return to baseline $\text{SaO}_2$ at the end of each event. We compared visually and automatically analyzed samples of tracings of all patients showing a series of desaturation events with a return to baseline $\text{SaO}_2$ at the end of each event, for a total of 5 h and 39 min of recording time. In this instance, visual and automatic calculation of the ODI showed no significant differences (automatic ODI of 82.3 vs visual ODI of 86.4; not significant).

**DISCUSSION**

Our study demonstrates that ODI automatic analysis performed by the MESAM 4 device is not reliable when baseline $\text{SaO}_2$ levels drop during re-
cordings. This usually happens in patients with complicated OSAS (as described in the introduction) whose severe clinical condition and need for urgent treatment justifies an unattended sleep study, according the ASDA guidelines. In these patients, the clinical judgment on the severity of OSAS and the following therapeutic decisions could be seriously misled if based on automatic ODI alone.

Some discrepancies between automatic and manual calculation of ODI have been mentioned in previous articles but were not systematically analyzed. In their validation study, Stoohs and Guillemi-
nault\textsuperscript{4} described three patients who presented major differences in the RDI calculated by polysomnography and the ODI calculated by the MESAM 4 device (RDIs of 83.5, 84.9, and 88.9 vs ODIs of 40, 38, and 51, respectively). The discrepancies were ascribed to the difference in the “total sleep time” and “total analysis time” calculation and to the fact that some events did not lead to \textit{SaO} \textsubscript{2} drops of > 3%.\textsuperscript{4} Given our findings, we wonder whether these discrepancies could have been due to a fall in baseline \textit{SaO} \textsubscript{2} during sleep. Moreover, Koziej et al\textsuperscript{5} and Esnaola et al\textsuperscript{6} showed that visual scoring of MESAM 4 tracings is more reliable than automatic scoring in screening OSAS patients, but they did not look for different results depending on the severity of OSAS.

Our study emphasizes the ASDA recommendation

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.pdf}
\caption{MESAM 4 recording of the same patient as described in Figure 1, in standard (top) and full-disclosure mode. In full-disclosure mode, the desaturation events automatically detected are underlined by the device. The analysis begun at 11:10 PM (23 10) (arrow). The system initially scored the events. As soon as the baseline \textit{SaO} \textsubscript{2} dropped at 11:16 PM (23 16), the system no longer identified the events, which were visually clearly detectable, as long as \textit{SaO} \textsubscript{2} did not return to baseline levels (11:43 PM [23 43]). The recorded parameters and abbreviations are described in Figure 1.}
\end{figure}
on the use of portable systems in the assessment of OSAS stating that “computer analysis of recording data are acceptable only if used as an aid to interpretation in conjunction with visual inspection of the entire raw data report”.\(^2\) In fact, our results indicated that hand calculation of ODI on MESAM 4 recordings is necessary when studying a patient with complicated OSAS whose oximetry tracing shows falls in baseline Sa\(_{O_2}\) associated with desaturation events related to apneas. However, we did not find significant differences between manual and automatic scoring of ODI when we considered samples of tracing with desaturation events characterized by a return to baseline values; this seems to suggest that the MESAM 4 automatic analysis could be reliable in the diagnosis of severe OSAS in patients without associated pathologic conditions.

In conclusion, our study suggests that although the MESAM 4 system can be useful for studying OSAS patients with severe clinical symptoms who need urgent sleep study in order to start treatment, the indexes obtained automatically should always be compared with visual evaluation of all-night recordings performed by personnel trained in polysomnography.

REFERENCES